Bayesian Hierarchical Modelling of Dual Response Surfaces

Younan Chen

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

Keying Ye, Chair
Angela N. Patterson
Samantha C. Prins
Eric P. Smith
G. Geoffrey Vining

November 29, 2005
Blacksburg, Virginia

Keywords: Bayesian Hierarchical Model, Dual Response Surfaces, Genetic Algorithm

Bayesian Hierarchical Modelling of Dual Response Surfaces

by

Younan Chen

Keying Ye, Chair

Department of Statistics

(ABSTRACT)

Dual response surface methodology (Vining and Myers [40]) has been successfully used as a cost-effective approach to improve the quality of products and processes since Taguchi [36] introduced the idea of robust parameter design on the quality improvement in the United States in mid-1980s. The original procedure is to use the mean and the standard deviation of the characteristic to form a dual response system in linear model structure, and to estimate the model coefficients using least squares methods.

In this dissertation, a Bayesian hierarchical approach is proposed to model the dual response system so that the inherent hierarchical variance structure of the response can be modelled naturally. The Bayesian model is developed for both univariate and multivariate dual response surfaces, and for both fully replicated and partially replicated dual response surface designs. To evaluate its performance, the Bayesian method has been compared with the original method under a wide range of scenarios, and it shows higher efficiency and more robustness. In applications, the Bayesian approach retains all the advantages provided by the original dual response surface modelling method. Moreover, the Bayesian analysis allows inference on the uncertainty of the model parameters, and thus can give practitioners complete information on the distribution of the characteristic of interest.

Dedication

To my parents, Zhiliang Chen and Qiu'er Zhao, who have always supported and encouraged me in my endeavors. The lessons they have taught me will last a lifetime.

Acknowledgements

I would like to thank my family, especially my parents and my husband, Bin Le, for their patience, love, support and encouragement. The love and support of my family has enabled me to make this dream true.

I would especially like to thank Dr. Keying Ye, who has been there for me every step of the way, teaching, encouraging and coaching, His patience and wisdom have been invaluable.

I would also express my appreciation to the committee members Dr. Angela N. Patterson, Dr. Samantha C. Prins, Dr. Eric P. Smith, and Dr. G. Geoffrey Vining for their valuable comments and suggestions.

Contents

1 Introduction **1**

2 A Bayesian Hierarchical Approach to Dual Response Surface Modelling **5**

2.1 Introduction 6

2.2 Least Square Methods and Bayesian Hierarchical Modelling 8

2.2.1 Review of the frequentist least square methods 8

2.2.2 Bayesian hierarchical modelling 9

2.3 The Optimization Algorithms 11

2.4 Bayesian Hierarchical Modelling in DRS 12

2.4.1 Notations and assumptions 12

2.4.2 Model building 13

2.4.3 Computation 15

2.5 Methods Comparison 16

2.5.1 Printing ink data 17

2.5.2 Catapult data 20

2.6 Simulation Studies 22

2.7 Summary and Conclusion 25
2.8 Appendix 26
2.8.1 Posterior distributions (conventional informative priors) 26
2.8.2 Posterior distributions (noninformative priors) 27
2.8.3 Propriety of the posterior distributions (noninformative priors) 28

3 Bayesian Hierarchical Modelling on Dual Response Surfaces in Partially Replicated Designs 31
3.1 Introduction 32
3.2 Application of the Genetic Algorithm in DRS and Optimization of the Quality and Reliability Measure 34
3.2.1 The genetic algorithm and hybrid computation methods 34
3.2.2 Optimization of the quality and reliability measure 35
3.3 Bayesian Hierarchical Modelling in DRS 37
3.3.1 The Bayesian hierarchical model 37
3.3.2 Extension of the Bayesian hierarchical model to partially replicated designs 39
3.4 An Example 41
3.5 Simulation Studies 45
3.6 Summary and Conclusion 51
3.7 Appendix 52
3.7.1 Posterior distributions (conventional informative priors) 52
3.7.2 Posterior distributions (noninformative priors) 54
3.7.3 Propriety of the posterior distributions (noninformative priors) 55

4 Assessment of the Bayesian Hierarchical Approach in Modelling Dual Response Surfaces **57**

4.1 Introduction . . . 58

4.2 Commonly Used Second Order Designs . . . 59

4.2.1 Central composite designs . . . 59

4.2.2 Small composite designs . . . 60

4.2.3 Hoke designs . . . 60

4.2.4 Notz designs . . . 62

4.3 Simulation Studies . . . 62

4.3.1 Fixed parameters, $k = 3$, normal distributions . . . 63

4.3.2 Fixed parameters, $k = 4$, normal distributions . . . 75

4.3.3 Fixed parameters, $k = 3$, t_4 distribution . . . 75

4.3.4 Random parameters, $k = 3$, normal distributions . . . 76

4.4 Conclusions . . . 82

5 Bayesian Methodology for Multivariate Dual Response Surfaces **83**

5.1 Introduction . . . 83

5.2 Extension of the Bayesian Hierarchical Model . . . 85

5.3 An Example . . . 85

5.4 Comparison of Modelling Efficiencies Under Covariance Matrix Uncertainty . . . 89

6 Summary and Future Research **93**

List of Figures

3.1 Boxplots for the logarithm of the relative efficiency ratios when the factorial portion is replicated and the variance model contains linear terms. 48

3.2 Boxplots for the logarithm of the relative efficiency ratios when the factorial portion is replicated and the variance model contains linear and interaction terms. 49

3.3 Boxplots for the logarithm of the relative efficiency ratios when the axial portion is replicated and the variance model contains linear terms. 50

4.1 Design matrices for several $k = 3$ second order designs: (a) SCD, (b) Hoke $D4$, and (c) Notz design. 61

4.2 Boxplots for the logarithm of the relative efficiency ratios when the CCD is fully replicated and the variance model is of complete second order. Model parameters are simulated from multivariate normal distributions. 78

4.3 Boxplots for the logarithm of the relative efficiency ratios when the CCD is factorial replicated and the variance model contains linear terms. Model parameters are simulated from multivariate normal distributions. 79

4.4 Boxplots for the logarithm of the relative efficiency ratios when the CCD is factorial replicated and the variance model contains linear and interaction terms. Model parameters are simulated from multivariate normal distributions. 80

4.5 Boxplots for the logarithm of the relative efficiency ratios when the CCD is axial replicated and the variance model contains linear terms. Model parameters are simulated from multivariate normal distributions. 81

List of Tables

2.1 Printing ink data 18

2.2 Optimization results comparison for the printing ink data 19

2.3 Catapult data 21

2.4 Optimization results comparison for the catapult data 22

2.5 Comparison of the relative efficiencies between the three methods using simulated data. Simulation size is 2500 for each scenario. It is determined so that the largest standard error of the mean relative efficiency is less than 0.005. . . 24

3.1 Optimization results comparison for the printing ink data 43

3.2 Using AGA to optimize the quality and reliability measure 45

3.3 The relative efficiencies of the three methods in the case of replicated factorial designs and linear variance models. Simulation size is 2500 for each scenario. 47

3.4 The relative efficiencies of the three methods in the case of replicated factorial designs and linear plus interaction variance models. Simulation size is 2500 for each scenario. 49

3.5 The relative efficiencies of the three methods in the case of replicated axial designs and linear variance models. Simulation size is 2500 for each scenario. 50

4.1 The relative efficiencies of the three methods when the basic design is a 3^3 complete cubic design. 65

4.2 The relative efficiencies of the three methods when the basic design is a central composite design with the number of the control factors $k = 3$ and the axial distance $\alpha = 1.682$. 66

4.3 The relative efficiencies of the three methods when the basic design is a small composite design with the number of the control factors $k = 3$ and the axial distance $\alpha = 1.682$. 67

4.4 The relative efficiencies of the three methods when the basic design is a Hoke design with the number of the control factors $k = 3$. 68

4.5 The relative efficiencies of the three methods when the basic design is a Notz design with the number of the control factors $k = 3$. 69

4.6 The relative efficiencies of the three methods when the basic design is a central composite design with the number of the control factors $k = 4$ and the axial distance $\alpha = 2$. 70

4.7 The relative efficiencies of the three methods when the response comes from a t_4 distribution, and the basic design is a central composite design with the number of the control factors $k = 3$ and the axial distance $\alpha = 1.682$. 71

5.1 Partial data from the chemical vapor deposition experiment 86

5.2 Comparison of the relative efficiencies under different correlations. The simulation size is determined such that the larger standard error of the relative efficiencies for the two methods does not exceed 0.01. 91

Chapter 1

Introduction

Introduced by Box and Wilson ([5]), response surface methodology (RSM) is a collection of statistical design techniques, empirical model-building procedures, and optimization methods that are intended for use in attaining optimal operating conditions for processes. Much of RSM, particularly in the early years, is focused on finding operating conditions that resulted in an optimum of the mean response with the homogeneity assumption on the variances.

However, the quality of a process or a product is not characterized solely by the mean of a defined characteristic. In off-line quality control, the variability of the characteristic is an important aspect to be monitored apart from the mean. Taguchi ([36]) first points out the need for monitoring both the mean and variance of an industrial process. In Taguchi's philosophy, there are two types of factors in an industrial system (a product or a process): factors that cannot be easily controlled in the manufacturing stage, though they can be in the experiment stage, and factors that can be easily controlled in both the experiment and the manufacturing. Taguchi refers to those not-easy-to-control factors as noise factors and easy-to-control factors as control factors. The variability of a defined characteristic consists of natural variation and variation introduced by noise variables. To reduce the variation from noise variables, the experimenter needs to find the appropriate levels of control factors at which the process or the product is insensitive to the change of the noise factors. The optimization task is to find the operating conditions where the mean response achieves the pre-specified target value and the variability around the target value is small. Such a technique is called *robust parameter design*. Taguchi's philosophy catalyzes the development of *robust parameter design* since early 1980's. However, he develops and adopts a package of tools (such as crossed-array designs and signal-to-noise ratios), which has been widely criticized for lack of statistical foundation. Nair *et al.* ([28]) provide a complete panel discussion on his approach.

The dual response surface (DRS) methodology is proposed by Vining and Myers ([40]) as a more rigorous statistical method within the response surface methodology framework to achieve Taguchi's philosophy. As with other optimization problems in RSM, there are also three main stages in DRS methodology. The first stage is to build a properly designed experiment so that the data containing the information among the responses and the control factors can be obtained efficiently and parsimoniously. The second stage is to explore the relationship between the control factors and the responses. In this stage, two models are built based on the data from the experiment, one for the process characteristic and the other for the process variance. The last stage is to search for the optimal operating condition throughout the region of interest under certain optimization criterion based on the two established models. The third stage results cannot be trusted if models built in the second stage do not reflect the dual response surfaces well. The second stage is the focus of our research, and all work is done under the assumption that data have already been collected. Modelling efficiency is assessed by comparing the performance of a product or a process at the optimal operating conditions predicted by each method, therefore optimization criteria and algorithms are inevitably involved in our work.

Following Vining and Myers ([40]), several optimization formulations and procedures have been proposed for DRS problems, such as Del Castillo and Montgomery ([12]), Lin and Tu ([24]) and Tang and Xu ([37]). The above optimization work is confined to the third stage and proceeds under the assumption that the established models approximate the true response surfaces well. Therefore the accuracy of the optimization results largely relies on whether the two established models are good approximations of the true dual response surfaces or not. The quality of the approximation mainly depends on the complexity of the true dual response surfaces, the effectiveness of the design strategy, and the efficiency of the modelling method. The complexity of the true dual response surfaces is determined by the nature of the characteristic of interest, which we have no way to exert influence upon. The effectiveness of the design strategy is an issue which should be addressed before the data is obtained. Therefore once a data set has been obtained and we want to get a good approximation of the dual response surfaces, we need to have a systematic modelling approach which is overall efficient and robust to various dual response surfaces and design strategies.

Up to now all modelling methods in the second stage are suggested within the frequentist-statistics scheme. The primary objective of this research is to develop a Bayesian hierarchical approach to DRS modelling based on the inherent hierarchical structure within DRS problems. After the model development, various theoretical and computational issues associated with the Bayesian model are addressed in details. The modelling performance of the proposed approach is evaluated against the frequentist models under various scenarios. The Bayesian model consistently displays higher efficiency and robustness.

In practice, the Bayesian model retains the advantages of the dual response approach: it has two separate estimated models to monitor both the mean characteristic and the process variability. Another advantage of the Bayesian approach is that it can incorporate prior information from previous studies into the current data analysis, and thus it provides flexibility in modelling when prior knowledge of the parameters is available. Moreover, as the Bayesian methodology works with the probability distributions, it facilitates the inference on the predictive distribution of the response at any point within the region of interest. Hence analysts will be given more comprehensive views of the response of interest rather than merely the mean and the variance.

The dissertation is organized as follows. Chapter 2 proposes a Bayesian hierarchical approach to model dual response surfaces, and applies it to fully replicated dual response surfaces. In that chapter, the Genetic Algorithm (De Jong [10]) is applied as the first step in optimization, followed by a local optimization algorithm. Chapter 3 extends the application of the Bayesian hierarchical model to partially replicated designs. Moreover, that chapter explores another advantage of Bayesian analysis: to use the posterior distributions of the model parameters to draw inference on the reliability of a predicted optimal condition. An adaptive version of the Genetic Algorithm is designed to optimize the reliability. Chapter 4 provides complete assessments on the modelling efficiencies of the least squares methods and the Bayesian approach under various scenarios. In Chapter 5, the Bayesian hierarchical approach is extended to model multivariate dual response surfaces, and is compared with frequentist methods in term of modelling efficiencies. Future research on the Bayesian hierarchical model for dual response surfaces is sketched in Chapter 6.

Chapter 2

A Bayesian Hierarchical Approach to Dual Response Surface Modelling

Abstract

In modern quality engineering, dual response surface methodology is a powerful tool to monitor an industrial process by using both the mean and the standard deviation of the measurements as the responses. The least squares method in regression is often used to estimate the coefficients in the mean and standard deviation models, and various decision criteria are proposed by researchers to find the optimal conditions. Based on the inherent hierarchical structure of the dual response problems, we propose a hierarchical Bayesian approach to model dual response surfaces. Such an approach provides more flexibility in modelling. This approach is further compared with two frequentist least squares methods by using two real data sets and simulated data.

Key Words: Bayesian hierarchical model; off-line quality control; genetic algorithm; optimization.

2.1 Introduction

Much of response surface methodology (RSM), particularly in the early years, was focused on finding operating conditions that resulted in an optimum of the mean response assuming homogeneous variances. During the last two decades, industrial statisticians and practitioners have become aware that they can no longer focus themselves only on the expected value of the response of interest. Instead, the variability of the response also needs to be considered. A common problem in an industrial process is to find the operating condition that achieves the target value for the mean of a process characteristic and minimizes the process variability. The pioneering work has been credited to Taguchi ([36]), who developed a package of tools which were viewed unfavorably by many researchers and practitioners for lack of statistical foundation. See Nair *et al.* ([28]) for detailed discussions.

The dual response surface approach, first introduced by Myers and Carter ([27]) and revitalized by Vining and Myers ([40]), suggests that the process characteristic and its process variability form a dual response system (DRS), and two separate models are established for the response and its variance. Almost in the same period, Welch *et al.* ([43]) and Shoemaker *et al.* ([34]) suggest using two separate models for robust design studies. In statistics, this approach allows the use of all regression tools to approximate the two response surfaces. In practice, the two separate models give the analyst a more scientific understanding of the total process, and thus allow them to see what levels of the control factors can lead to satisfactory values of the response as well as the variance.

Like other optimization work in RSM, the dual response optimization problem also consists of the following three stages. The first stage is to build an optimal experiment so that the information among the responses and the control factors can be obtained efficiently. The second stage is to build two models based on the data from the experiment, one for the process characteristic and the other for the process variance. The last stage is to search for the optimal operating condition throughout the region of interest under certain optimization criterion based on the established models. The third stage results cannot be trusted if models built in the second stage do not reflect the dual response surfaces well. The second stage is the focus of this paper, and all work is done under the assumption that data have already been collected. Model building efficiency is usually evaluated by comparing the performance of a product or a process at the found operating conditions, therefore optimization criteria and algorithms are inevitably involved.

Following Vining and Myers ([40]), several optimization formulations and procedures have been proposed for the DRS problem, for example, in Del Castillo and Montgomery ([12]), Lin and Tu ([24]), Copeland and Nelson ([9]), Ames et al. ([2]), Kim and Lin ([23]) and Tang and Xu ([37]). The above optimization work is confined to the third stage and is carried out under the assumption that the established models approximate the true response surfaces well. Therefore the accuracy of the optimization results largely relies on whether the two established models are good approximations of the true dual response surfaces or not. If the two estimated models fit the surfaces poorly, then the true response and its variance at the chosen operating condition are very likely to be far from the specified requirement.

In this paper, a Bayesian hierarchical regression modelling approach is proposed as an analysis tool for dual response surface problems. The sample means are used as the response for the mean model and log-normal distributions are assumed for the variances. The estimates of the coefficients in the two models are based on posterior inference. A hybrid of local optimization algorithms and the genetic algorithm is adopted to search for the optimal operating conditions under two common optimization criteria. This study not only gives an ad hoc Bayesian method for the analysis of dual response surfaces, but also provides more flexibility for modelling in situations where prior knowledge of the parameters is available.

In Section 2.2, the frequentist modelling approach initially proposed by Vining and Myers ([40]) is briefly reviewed and the basic idea of the Bayesian hierarchical model is sketched. Section 2.3 presents a brief introduction of the genetic algorithm and discusses its advantages and disadvantages. Furthermore, a hybrid optimization method is proposed. In Section 2.4, a Bayesian hierarchical model is developed and the associated computation issues are discussed. After the model development, the Bayesian approach is compared with the two least squares methods by using two real data sets and simulated data. The theoretical details of the Bayesian approach are placed in the Appendix 2.8.

2.2 Least Square Methods and Bayesian Hierarchical Modelling

2.2.1 Review of the frequentist least square methods

Let $\underline{x}$ represent a $k \times 1$ vector of independent factors under the experimenter's control and $\mathbf{X}$ an $N \times p$ matrix used in the model building, where N is the number of different design locations and p the number of coefficients in the model. If the model is of complete second-order, the matrix $\mathbf{X}$ consists of the k coordinates of the design points plus the intercept, the interaction terms and the quadratic terms. Suppose that the experimenter seeks to optimize the process for $\underline{x} \in R$ where R is the region of interest (usually the experimental region). Vining and Myers ([40]) build two full second order models for the sample mean and the sample standard deviation respectively:

$$\underline{\overline{y}} = \mathbf{X}\underline{\beta} + \varepsilon, \quad \text{and} \quad \underline{s} = \mathbf{X}\underline{\gamma} + \eta, \tag{2.1}$$

where $\underline{\overline{y}}$ is the vector of the sample means, $\underline{s}$ the vector of the sample standard deviations at the design points, $\underline{\beta}$ and $\underline{\gamma}$ are the coefficient vectors to be estimated, and ε and η are the error terms in their respective model.

The least squares method is a natural choice to estimate the coefficients for the dual response models without any assumption on the distributions of the sample mean or the sample standard deviation. Vining and Myers ([40]) mention that the generalized least squares (GLS) method should be pursued to estimate $\underline{\beta}$ in order to take into account the heterogeneity of variances in dual response problems:

$$\underline{\widehat{\beta}} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\underline{\overline{y}} \quad \text{and} \quad \underline{\widehat{\gamma}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\underline{s},$$

where $\hat{\mathbf{V}}$ is the estimated covariance matrix of the mean responses obtained in the design. $\hat{\mathbf{V}}$ is diagonal with the assumption that the random errors are independent from design point to design point. Due to computational difficulty, only ordinary least squares estimates were calculated in their paper. The diagonal elements of $\hat{\mathbf{V}}$ can be obtained in two ways. One is to use the sample variance at each design point. The other way is to use predicted values obtained through the variance model, by which the information in the process variance model can be incorporated into the mean response model. When the number of replicates is small, the latter one is usually preferred. On the other hand, the iterative least squares method is also viewed as a possible approach to improve the generalized least squares estimates. However, through simulation studies the iterative method is found to be of similar modelling performance as the generalized least squares.

One pitfall with the modelling approach in (2.1) is that it is possible to yield negative predicted values of the standard deviation, even if the true mean standard deviations are positive throughout the region of interest. If the response at a point with the negative predicted standard deviation happened to be the optimum under certain criterion, it would result in difficulty in explaining the process performance at the picked optimal point, and cause confusion among the practitioners. Hence, we are trying to seek a more natural modelling mechanism that would fit the dual response system well.

2.2.2 Bayesian hierarchical modelling

Bayesian hierarchical models mainly deal with data involving multiple parameters that can be treated as related through the structure of the data. Suppose there are observed data y_{ij}'s, where y_{ij} is the j^{th} observation in the i^{th} group. The grouping of the data is determined by different settings of the study, such as locations, times and environmental factors. The responses within groups are more homogeneous than between groups. The vector $\underline{\theta}_i$ (could be a scalar) contains the parameter(s) to determine the distribution of y_{ij} in the i^{th} group, and they are believed to be connected with parameters $\underline{\phi}$. The parameters $\underline{\phi}$ are at a higher level and are called *hyperparameters*. Conditional on $\underline{\theta}_i$, the data y_{ij}'s are assumed to be independently distributed, and given the hyperparameters $\underline{\phi}$, the parameters $\underline{\theta}_i$'s have the common density $\pi(\underline{\theta}|\underline{\phi})$.

Nonhierarchical models are usually inappropriate for a data set with such a hierarchical structure. If all y_{ij}'s are assumed to come from the same population distribution, the difference among the groups is inevitably neglected. If the uniqueness of each group needs to be introduced into the model without using a hierarchical model, too many parameters have to be entered into the model and the data will be over-fitted. As a result, the estimated model may fit the existing data well but will produce poor predictions for future observations.

What a Bayesian hierarchical model does is to express the relationship among the parameters $\underline{\theta}_i$'s with a prior distribution: $\underline{\theta}_i$'s are treated as variables and $\underline{\phi}$ as the parameters in the prior distribution of the $\underline{\theta}_i$'s. In this way, the only unknown parameters in the model are $\underline{\phi}$. Denote by $\pi(\underline{\theta}, \underline{\phi})$ the joint prior distribution of $\underline{\theta}$ and $\underline{\phi}$, $\pi(\underline{\theta}|\underline{\phi})$ the conditional prior distribution of $\underline{\theta}$ given $\underline{\phi}$, and $\pi(\underline{\phi})$ the prior distribution of $\underline{\phi}$. Furthermore, let $f(\text{data}|\underline{\theta})$ represent the probability density function of the data given $\underline{\theta}$, and $\pi(\underline{\theta}, \underline{\phi}|\text{data})$ the posterior distribution of the parameters. The joint prior distribution of parameters can be expressed as:

$$\pi(\underline{\theta}, \underline{\phi}) = \pi(\underline{\theta}|\underline{\phi}) \cdot \pi(\underline{\phi}),$$

and the posterior distribution is:

$$\pi(\underline{\theta}, \underline{\phi}|\text{data}) \propto f(\text{data}|\underline{\theta}, \underline{\phi}) \cdot \pi(\underline{\theta}, \underline{\phi}) = f(\text{data}|\underline{\theta}) \cdot \pi(\underline{\theta}|\underline{\phi}) \cdot \pi(\underline{\phi}),$$

since $f(\text{data}|\underline{\theta}, \underline{\phi})$ depends only on $\underline{\theta}$, and the hyperparameters $\underline{\phi}$ affect the data only through $\underline{\theta}$. The prior distribution facilitates the incorporation of subjective information into the analysis of the current data set. If little is known about $\underline{\phi}$, a noninformative prior distribution can be specified followed by a check on the propriety of the joint posterior distribution. More discussions of the Bayesian hierarchical modelling can be found in Gelman *et al.* ([16]).

Bayesian analysis is based on the posterior inference. Estimates of parameters are usually summary statistics of the marginal posterior distributions, such as the posterior mean, median, mode, standard deviation, etc. In simple nonhierarchical Bayesian models, it is often easy to analytically derive the marginal posterior distributions and obtain the summary statistics. However, in more complicated models, especially if the parameters are of multiple dimensions, it is often hard or impossible to analytically express the marginal distribution of each parameter. A common strategy is to use simulation techniques, such as MCMC Gibbs sampling, to generate observations for the marginal posterior distributions and calculate the summary statistics numerically. See Robert and Casella ([33], chap. 7) for more information.

Bayesian hierarchical modelling of dual response system is a natural application as the hierarchical structure is a built-in property of dual response surface problems: the characteristic measurements of the process obtained at the same design setting share more common properties than those obtained among different settings; the parameters in the mean response model, including the variance, determine the distribution of the process characteristic at each design point; and the distribution of that variance can be further determined by the coefficients in the process variance model.

2.3 The Optimization Algorithms

In this paper, we use the combination of the genetic algorithm (GA) and local optimization algorithms, such as the Generalized Reduced Gradient (GRG) method and Sequential Quadratic Programming (SQP) methods to achieve dual response optimization. See Del Castillo and Montgomery ([12]) and Fletcher and Powell ([15]) for more information about GRG and SQP, respectively.

The genetic algorithm is pioneered by Holland ([19]), and applications of GA to problems of mathematical optimizations owe much to De Jong ([10]). Recently it has emerged as an increasingly popular family of methods for global optimization, mainly because it offers a simple way (for both continuous and discrete variables) to solve for near-global optimum even for poorly behaved functions. The idea of GA is to apply the principles of "Darwinian natural selection" iteratively to a population of computer representations of the solution domain. The algorithm attempts to mimic the natural evolution of a population by allowing solutions to reproduce, to create new solutions, and to compete for survival in the next iteration. After many generations (iterations), the best solution is usually near the global optimum.

Instead of the point-to-point search in the traditional methods, GA proceeds from one population to another, and thus it sweeps through the parameter space in many directions simultaneously and thereby reduces the probability of convergence to false optima. The optimal locations resulting from GA are always close to each even for poorly behaved surfaces.

The main limitation of GA is that, like other direct-search methods, convergence of the genetic algorithm does not necessarily occur at a point where the gradient is exactly 0. That is, the solution from GA does not guarantee the exact global optimum, but a near global optimum. If a very precise result is required, GA may suffer from excessively slow convergence because of its fundamental requirement of neglecting the local information. However, GA will typically find a point that is close enough to the optimum so that a gradient-type algorithm will efficiently converge to the exact global optimum if the found location is used as the starting point.

On the other hand, classical "hill-climbing" methods are well known to exploit all local information in an efficient way, yet they often result in different local optima if different starting points are given. To take the advantages of both types of the algorithms, we use a hybrid method for optimization: GA is first used to find a near-global optimal point, and then the chosen point serves as the starting point for the local optimization algorithm to identify the exact global optimum. In this paper, the genetic algorithm is designed and developed using Matlab specifically for dual response problems. The subsequent local optimization is conducted with SQP methods using a Matlab built-in function.

2.4 Bayesian Hierarchical Modelling in DRS

The Bayesian hierarchical model is proposed as a new approach to dual response surface modelling. The coefficients in the dual models can be estimated using location parameters of the posterior distributions. Consequently, the predicted mean and variance can be obtained at any point within the design region.

2.4.1 Notations and assumptions

Let $\underline{x}$ represent a $k \times 1$ vector of independent control factors, $\mathbf{X}$ an $N \times p$ matrix used in the mean response model and $\mathbf{Z}$ an $N \times q$ matrix in the variance model, where N is the number of distinct design points in the experiment, p and q are the numbers of parameters in the mean and variance models respectively; and $\underline{x}_i'$, $\underline{z}_i'$ are the respective row vectors associated with the matrices $\mathbf{X}$ and $\mathbf{Z}$ for the i^{th} location in the experiment. Denote by $\mathbf{Y} = (y_{11}, y_{12}, \ldots, y_{1n_1}, y_{21}, \ldots, y_{Nn_N})'$ the vector of observations from the experiment, and y_{ij} the j^{th} response observed at the i^{th} design point, where $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, n_i$; and the sample variance and population variance at the i^{th} design point are denoted by s_i^2 and σ_i^2 respectively. Furthermore, for all the parameters, $\underline{\beta}$ represents a $p \times 1$ vector of coefficients of the mean response model, $\underline{\gamma}$ a $q \times 1$ vector of coefficients of the variance response model, and δ^2 the hyperparameter in the distribution of σ_i^2's.

Note that $\mathbf{X}$ and $\mathbf{Z}$ are not necessarily the same, and quite often q is less than p as variance models are usually simpler than mean response models.

We further assume independence among observations and parameters as follows:

- Independence among y_{ij}'s conditional on $\underline{\beta}$ and σ_i^2's:

$$corr(y_{ij}, y_{i'j'} | \underline{\beta}, \sigma_i^2, \sigma_{i'}^2) = 0, \quad \text{for either} \quad i \neq i' \quad \text{or} \quad j \neq j' \quad \text{or both.}$$

- Independence among σ_i^2's conditional on $\underline{\gamma}$ and δ^2:

$$corr(\sigma_i^2, \sigma_{i'}^2 | \underline{\gamma}, \delta^2) = 0, \quad \text{for} \quad i \neq i'.$$

- Independence among $\underline{\beta}$, $\underline{\gamma}$, and δ^2 in the prior specification:

$$\pi(\underline{\beta}, \underline{\gamma}, \delta^2) = \pi(\underline{\beta}) \cdot \pi(\underline{\gamma}) \cdot \pi(\delta^2).$$

The assumption of independence among parameters is made for simplicity, though it may need to be modified in the presence of subjective information.

2.4.2 Model building

The Bayesian hierarchical model is built in the following three stages:

- **Stage 1** Since variances are assumed to be different at various locations in dual response surface problems, a natural choice for the process characteristic is a heteroscedastic normal distribution conditional on $\underline{\beta}$ and the variance σ_i^2 at the i^{th} location:

$$y_{ij}|\underline{\beta}, \sigma_i^2 \sim N(\underline{x}_i'\underline{\beta}, \sigma_i^2), \quad -\infty < y_{ij} < \infty, \quad \text{for} \quad i = 1, 2, \ldots, N; \quad j = 1, 2, \ldots, n_i.$$

Hence the distribution of the sample mean at the i^{th} location is $N(\underline{x}_i'\underline{\beta}, \sigma_i^2/n_i)$.

- **Stage 2** The logarithm transformation is a common technique used in variance modelling. Bartlett and Kendall ([3]) use the logarithm of the sample variance as the response for a linear regression model, and Engel and Huele ([14]) use the logarithm of the residual squares. Similarly, in the Bayesian model, the variance σ_i^2 is assumed to have a log-normal distribution conditional on $\underline{\gamma}$ and δ^2:

$$\ln \sigma_i^2 | \underline{\gamma}, \delta^2 \sim N(\underline{z}_i'\underline{\gamma}, \delta^2) \quad \text{for} \quad i = 1, 2, \ldots, N.$$

- **Stage 3** Set the prior distributions for parameters, $\underline{\beta}$, $\underline{\gamma}$, and δ^2. Informative priors are always preferred if prior information with respect to parameters is available from previous studies or subjective knowledge. In choosing the form of the informative priors, those from the conjugate family are of the most interest. The practical advantage of using a conjugate prior distribution is that we know the posterior distribution follows the same parametric form as of the prior distribution. However, due to the complexity of the model we consider here, it is hard to find the conjugate priors. Therefore the commonly used prior distributions for such kind of parameters are assumed as follows:

$$\pi(\underline{\beta}) \sim \text{MVN}_p(\underline{\mu}_\beta, \Sigma_\beta), \quad \pi(\underline{\gamma}) \sim \text{MVN}_q(\underline{\mu}_\gamma, \Sigma_\gamma), \quad \pi(\delta^2) \sim \text{Inverse-Gamma}(\alpha, \tau),$$

with $\pi(\delta^2) = \frac{\tau^\alpha}{\Gamma(\alpha)}(\delta^2)^{-(\alpha+1)} \exp(-\tau/\delta^2)$, where $\underline{\mu}_\beta$, Σ_β, and $\underline{\mu}_\gamma$, Σ_γ are the respective mean vectors and covariance matrices of $\underline{\beta}$ and $\underline{\gamma}$ in the prior distributions, and both α, $\tau > 0$, are the shape and scale parameters of the Inverse-Gamma distribution, respectively.

If available information is in the form apparently not belonging to the conjugate family, we need to elicit more realistic prior distributions to incorporate such information. However, if there is no subjective information, noninformative priors could be used.

As discussed in Section 2.2.2, if the marginal posterior distributions of the parameters of interest cannot be analytically expressed, simulation techniques such as the MCMC Gibbs sampling method may be used. The joint posterior distribution and full conditional posterior distributions for using the above conventional prior distributions and noninformative distributions are needed to use the Gibbs sampling technique, and derived in Appendix 2.8.1.

For a dual response surface problem without any subjective information on parameters, the following independent prior distributions can be used:

$$\pi(\underline{\beta}) \propto 1 \quad \pi(\underline{\gamma}) \propto 1 \quad \pi(\delta^2) \propto \frac{1}{\delta^2} \exp(-\frac{\lambda}{\delta^2}),$$

where λ is a small positive number. The joint and conditional posterior distributions using the above noninformative priors are derived in Appendix 2.8.2. The propriety of the posterior distribution by using these improper priors are discussed in Appendix 2.8.3.

2.4.3 Computation

The Gibbs sampler is used in the posterior simulation. Based on Theorem 2.2 and Corollary 2.2, the variates of the full posterior conditional distributions of $\underline{\beta}$, $\underline{\gamma}$ and δ^2 can be generated directly but the distributions of σ_i^2's are uncommon, so the rejection method is used. To simulate observations for the desired density $f(x)$ with the rejection method, we have two requirements: first, we need a well-known density $g(x)$ that we can easily generate observations from; and second, the ratio between $f(x)$ and $g(x)$ must have an upper bound M. The basic rejection method is to generate an observation from $g(x)$ first, and then compare the value $\frac{f(x)}{Mg(x)}$ with a uniform variate ν, where $\nu \sim U(0,1)$, to decide whether the observation is accepted or not. The desired density $f(x)$ is often called the *target density*, and $g(x)$ is called the *instrumental density*. See Robert and Casella (1985, pp. 49-53) for more details.

The target density to be simulated is $\pi(\sigma_i^2|\text{others})$ in Corollary 2.2 and the instrumental density proposed is:

$$g(\sigma_i^2) \sim \text{Inverse-Gamma}\left(\frac{n_i}{2}, \frac{(n_i-1)s_i^2 + n_i(\bar{y}_i - \underline{x}_i'\underline{\beta})^2}{2}\right).$$

The ratio between the target density and the instrumental density is

$$R_i = \frac{\pi(\sigma_i^2|\text{others})}{g(\sigma_i^2)} \propto \exp\left[-\frac{1}{2\delta^2}(\ln\sigma_i^2 - \underline{z}_i'\underline{\gamma})^2\right] \quad \text{for} \quad i = 1, 2, \ldots, N.$$

When $\sigma_i^2 = \exp\{\underline{z}_i'\underline{\gamma}\}$, the ratio R_i achieves its maximum, which is the ratio between the normalizing constants of the target density and the instrumental density.

It is expected that adding a small positive value of λ should exert no influence on the conditional posterior distributions of $\underline{\beta}$, $\underline{\gamma}$, δ^2 and σ_i^2's. Usually $\frac{1}{2}(\underline{d} - Z\underline{\gamma})'(\underline{d} - Z\underline{\gamma})$ is very large relative to λ if λ is chosen to be tiny, so the conditional posterior distribution of δ^2 is little affected, and the whole simulation procedure remains almost unchanged. The printing ink data and the catapult data in Section 2.5 are used for sensitivity analysis to study the effect of changing λ on the estimated optimal locations. In choosing a small positive value for λ, we consider 0.1 as the largest small value we could use. Therefore $\lambda = 0.1, 0.001$, and 10^{-6} are used, and it is found that the predicted performance at the chosen locations is fairly robust to the choice of λ (refer to Tables 2.2 and 2.4 in Sections 2.5.1 and 2.5.2, respectively).

2.5 Methods Comparison

In this section, the proposed Bayesian hierarchical model (henceforth referred as BAYES) is applied to two data sets. The coefficients in the two models, $\underline{\beta}$ and $\underline{\gamma}$, are estimated with their respective posterior medians. Posterior medians are preferred over the posterior means for their robustness to outliers. The optimization results from the Bayesian model are compared with the two frequentist least squares methods in Vining and Myers ([40]): one is to model the mean and standard deviation with linear regressions and estimate both sets of parameters with ordinary least squares (OOLS); and the other is similar to OOLS except that the coefficients of the mean response models are estimated with generalized least squares (GOLS). Furthermore, simulation is used to assess the modelling efficiency of each method by comparing the performance at the chosen optimal operating conditions.

To compare the optimization result from each model, optimal operating conditions are found by searching under two criteria in the literature. The first one, proposed by Vining and Myers ([40]), strictly sets the mean response at the target value, and then minimizes the variability subject to this constraint (henceforth referred to as "target is the best" criterion). The second one, proposed by Lin and Tu ([24]), minimizes the sum of deviance around the target value and the variability. This measure is defined as (mean − target)^2 + variance, which is very similar to the definition of the mean squared error (MSE criterion).

2.5.1 Printing ink data

Box and Draper ([4]) outline an experiment involving printing ink, which has been widely adopted in dual response surface analysis for illustration, e.g. Ames et al. ([2]), Copeland and Nelson ([9]), and Kim and Lin ([23]). The purpose of the experiment was to study the effects of speed (x_1), pressure (x_2), and distance (x_3), upon a printing machine's ability to apply coloring inks upon package labels (y). The original experiment used a 3^3 complete factorial design with three runs at each design point for a total of 81 runs. All previous analysis considered the situation where we wish to minimize the process variability around a target value of 500 for the response. The same situation is considered in this paper. Since the experimental region is cuboidal with each factor taking values at $-1, 0, 1$, optimal conditions are searched for throughout the cuboidal region. The data is shown in Table 2.1.

Vining and Myers ([40]) assume full second order models for both the mean and variance models so that the dual response optimization algorithm in Myers and Carter ([27]) could be carried out. All subsequent analysis on this data set has been done based on this assumption, though some authors have pointed out that the assumption of two full second order models does not necessarily hold, e.g. Lin and Tu ([24]). In this paper, we will still stick to this assumption, i.e., the model matrices $\mathbf{X}$ and $\mathbf{Z}$ are identical, both containing columns for the intercept, linear terms, quadratic terms and cross product terms in order.

Optimization results from the three models are displayed in Table 2.2, in which $\underline{x}_a$ denotes the optimal location found under the "target is the best" criterion. $\hat{\sigma}_a^2$ and $\widehat{MSE}_a$ are respectively the estimated mean, variance and MSE at the location $\underline{x}_a$. Similarly,

Table 2.1: Printing ink data

Run	x_1	x_2	x_3	y_1	y_2	y_3	$\overline{y}$	s
1	-1	-1	-1	34	10	28	24	12.49
2	0	-1	-1	115	116	130	120.33	8.39
3	1	-1	-1	192	186	263	213.67	42.83
4	-1	0	-1	82	88	88	86.00	3.46
5	0	0	-1	44	178	188	136.67	80.41
6	1	0	-1	322	350	350	340.67	16.17
7	-1	1	-1	141	110	86	112.33	27.57
8	0	1	-1	259	251	259	256.33	4.62
9	1	1	-1	290	280	245	271.67	23.63
10	-1	-1	0	81	81	81	81.00	0
11	0	-1	0	90	122	93	101.67	17.67
12	1	-1	0	319	376	376	357.00	32.91
13	-1	0	0	180	180	154	171.33	15.01
14	0	0	0	372	372	372	372.00	0
15	1	0	0	541	568	396	501.67	92.50
16	-1	1	0	288	192	312	264.00	63.50
17	0	1	0	432	336	513	427.00	88.61
18	1	1	0	713	725	754	730.67	21.08
19	-1	-1	1	364	99	199	220.67	133.82
20	0	-1	1	232	221	266	239.67	23.46
21	1	-1	1	408	415	443	422.00	18.52
22	-1	0	1	182	233	182	199.00	29.44
23	0	0	1	507	515	434	485.33	44.64
24	1	0	1	846	535	640	673.67	158.21
25	-1	1	1	236	126	168	176.67	55.51
26	0	1	1	660	440	403	501.00	138.94
27	1	1	1	878	991	1161	1010.0	142.45

Table 2.2: Optimization results comparison for the printing ink data

Models	OOLS	GOLS	BAYES		
			$\lambda = 10^{-1}$	$\lambda = 10^{-3}$	$\lambda = 10^{-6}$
$\underline{x}_a$	(1.0000, 0.1163, -0.2584)	(1.0000, 0.7189 -0.4484)	(0.4788, 0.1695, 0.1002)	(0.4769, 0.1729, 0.0984)	(0.4757, 0.1759, 0.0975)
$\hat{\sigma}_a^2 = \widehat{MSE}_a$	2034.4	2494.7	27.824	28.472	28.536
$\underline{x}_b$	(1.000, 0.0715, -0.2503)	(1.0000, 0.5537, -0.3864)	(0.4783, 0.1695, 0.0998)	(0.4765, 0.1726, 0.0981)	(0.4753, 0.1757, 0.0972)
$\hat{\mu}_b$	494.672	493.036	499.845	499.833	499.840
$\hat{\sigma}_b^2$	1977.6	2421.6	27.783	28.416	28.483
$\widehat{MSE}_b$	2005.9	2470.0	27.808	28.444	28.508

notations $\hat{\mu}_b$, $\hat{\sigma}_b^2$ and $\widehat{MSE}_b$ are for the results found under the MSE criterion. These subscripts will also be used for in Table 2.4.

Though the only difference between the OOLS and GOLS models is in the estimation of β, the optimal operating conditions picked using the two methods are not close to each other and the predicted means, variances, and MSEs also differ greatly. This may cause confusion among the practitioners on which point should be viewed as "optimal". On the other hand, the optimal locations picked out using the Bayesian hierarchical methods for different small values of λ are very close to each other, and their respective predicted performances (such as predicted means, variances and MSEs) are quite similar. The potential scale reduction factor, well known as the Gelman-Rubin statistic, has been evaluated for all scalar estimates in the dual models for the ink data set, and they are all very close to one, which indicates a good convergence of the posterior distribution. See Gelman et al. ([16], pp. 332) for details. Similar convergence checking has been done in the analysis of the catapult data in Section 2.5.2, and it is found that the MCMC iterations converge well.

Table 2.2 shows a big discrepancy between the predicted variances and MSEs at the optimal conditions found by different methods. The variances of the Bayesian method are much smaller than those of the least squares methods. However, this does not come to a conclusion that Bayesian method is preferred against the other two methods as the nature of the data is unknown to us and the locations suggested by each model are only optimal to the extent that the fitted model is correct. One does not know whether the two least squares methods for this data set are underestimating the variances or the Bayesian method overestimates the variance. In Table 2.4 that follows, the opposite results will be shown for another data set. In conclusion, to compare the model efficiency of the three methods, a simulation study needs to be performed and they will be shown in Section 2.6.

2.5.2 Catapult data

The catapult data is from a Roman-style catapult experiment, first adopted by Luner ([25]) to illustrate, for teaching purposes, the use of RSM in quality improvement. The requirement for the catapult performance is that it must throw projectiles a distance of 80 inches with a high degree of precision. Based on the knowledge of the catapult production process, three factors, arm length (x_1), full stop angle (x_2), and pivot height(x_3), are included as potential significant factors to catapult performance. A completely randomized central composite design (CCD) is carried out: 8 factorial points and 6 axial points with 3 replicates at each point, and 18 replicates at the central point. The response y is defined as the distance, in inches, from the base of the catapult to the place where the projectile reaches the ground. Table 2.3 shows the results of the experiment.

As in the previous analysis, full second order models are assumed for both the mean and the variance in each method in the analysis of the catapult data. Optimal points, $\mathbf{x}_0$'s, are searched under the two optimization criteria used for the printing ink data over the spherical region, $\mathbf{x}_0'\mathbf{x}_0 \leq 3$, since the design region is almost spherical. Table 2.4 summarizes the optimization results from the three methods. For this data, the optimal points found using the OOLS and GOLS methods are not as different as those in the printing ink data set, and their predicted performances are much closer. Once again, the results from the BAYES method by using three different λ values are still very similar to each other. Furthermore,

Table 2.3: Catapult data

Run	x_1	x_2	x_3	y_1	y_2	y_3	$\bar{y}$	s
1	-1	-1	-1	39	34	42	38.33	4.04
2	-1	-1	1	80	71	91	80.67	10.12
3	-1	1	-1	52	44	45	47.00	4.36
4	-1	1	1	97	68	60	75.00	19.47
5	1	-1	-1	60	53	68	60.33	7.51
6	1	-1	1	113	104	127	114.67	11.59
7	1	1	-1	78	64	65	69.00	7.81
8	1	1	1	130	79	75	94.67	30.66
9	-1.682	0	0	59	51	60	56.67	4.93
10	1.682	0	0	115	102	117	111.33	8.14
11	0	-1.682	0	50	43	57	50.00	7.00
12	0	1.682	0	88	49	43	60.00	24.43
13	0	0	-1.682	54	50	60	54.67	5.03
14	0	0	1.682	122	109	119	116.67	6.81
15	0	0	0	87	78	89		
16	0	0	0	86	79	85		
17	0	0	0	85	81	87	84.89	4.06
18	0	0	0	89	82	87		
19	0	0	0	86	79	88		
20	0	0	0	88	79	90		

as discussed in Section 2.5.1, larger prediction variances and MSEs of the BAYES results do not mean that BAYES method produce worse estimation.

Table 2.4: Optimization results comparison for the catapult data

Models	OOLS	GOLS	BAYES		
			$\lambda = 10^{-1}$	$\lambda = 10^{-3}$	$\lambda = 10^{-6}$
$\underline{x}_a$	(0.0420, -0.2742, -0.2394)	(0.0435, -0.2829 -0.2227)	(-0.0616, -0.2285, -0.2302)	(-0.0573, -0.2177, -0.2345)	(-0.0618, -0.2427, -0.2263)
$\hat{\sigma}_a^2 = M\hat{S}E_a$	5.308	5.449	13.867	14.572	13.744
$\underline{x}_b$	(0.0595, -0.2741, -0.2633)	(0.0611, -0.2815, -0.2473)	(-0.0451, -0.1820, -0.2007)	(-0.0443, -0.1669, -0.2067)	(-0.0430, -0.1947, -0.1974)
$\hat{\mu}_b$	79.760	79.756	80.822	79.831	79.833
$\hat{\sigma}_b^2$	5.070	5.210	14.167	14.167	14.100
$M\hat{S}E_b$	5.127	5.270	14.842	14.195	14.128

2.6 Simulation Studies

As mentioned in Section 2.5.1, simulated data sets should be analyzed to assess the performance of each modelling approach.

We use the following three different scenarios to do the simulation:

- **Constant variance**:

$$y_{ij} \sim N(\underline{x}_i'\underline{\beta}, 1), \quad \text{for} \quad i = 1, 2, \ldots, N, \quad j = 1, 2, \ldots, n_j;$$

- Nonconstant variance where the standard deviation follows a normal distribution (referred to as **Nonconstant Stdev**):

$$y_{ij} \sim N(\underline{x}_i'\underline{\beta}, \sigma_i^2), \quad \sigma_i \sim N(\underline{z}_i'\underline{\gamma}, \eta^2), \quad \text{for} \quad i = 1, 2, \ldots, N, \quad j = 1, 2, \ldots, n_j.$$

In generating the standard deviation, the minimum mean standard deviation over the experimental region is set to be much larger than η, so that the simulated standard deviation has little likelihood to be negative;

- Nonconstant variance where the variance follows a log-normal distribution (referred to as **Nonconstant Logvar**):

$$y_{ij} \sim N(\underline{x}_i'\underline{\beta}, \sigma_i^2), \quad \sigma_i^2 \sim \text{Log-Normal}(\underline{z}_i'\underline{\gamma}, \eta^2), \quad \text{for} \quad i = 1, 2, \ldots, N, \quad j = 1, 2, \ldots, n_j.$$

Since full second order models are assumed for both the mean and variance models in previous analysis, $\underline{x}_i'$ and $\underline{z}_i'$ will be set as identical in the simulation, both containing the intercept, linear terms, quadratic terms and interactions in order. The GOLS estimates for the printing ink data are used as the parameters to simulate data in the Constant variance case and the Nonconstant Stdev case, and the Bayesian estimates are used for data generation in the Nonconstant Logvar case. For each scenario, 2500 data sets are generated from a 3^3 $(N = 27)$ factorial design with 3 replicates $(n_i = 3, \quad \text{for} \quad i = 1, 2, \ldots, 27)$ at each design point. The simulation size is determined such that the largest standard error of the mean efficiency is less than 0.005. Each method builds two full second order models for the mean and variance surfaces and optimal points are searched over the cuboidal region $(-1 < x_1, x_2, x_3 < 1)$, under the MSE criterion. For the BAYES model, $\lambda = 0.001$ is assumed in the prior distribution of δ^2 for the simulation studies as it has been shown that the effect of λ is almost negligible in the analysis for the two real data sets.

The relative efficiency based on the MSE criterion is introduced as a "single-number" measure of performance to see how well a model does in choosing "good" points overall. It is defined as the ratio between the true model's minimum MSE and the true model's MSE at the point proposed by that modelling approach:

$$\text{relative efficiency} = \frac{\min MSE_{\text{true}}}{MSE_{\text{point proposed}}}.$$

If the operating condition picked by a model performs closely to the true optimum, i.e., the relative efficiency is very near to 1, then the model is viewed as good. If the relative efficiency of a model is very small, that means the model picks a point far from optimum.

Table 2.5: Comparison of the relative efficiencies between the three methods using simulated data. Simulation size is 2500 for each scenario. It is determined so that the largest standard error of the mean relative efficiency is less than 0.005.

Models	**Constant Variance**			**Nonconstant Stdev**			**Nonconstant Logvar**		
	Mean	Median	Stdev	Mean	Median	Stdev	Mean	Median	Stdev
OOLS	0.9201	0.9561	0.0959	0.7714	0.8181	0.1693	0.5423	0.5500	0.2412
GOLS	0.9182	0.9552	0.0986	0.7695	0.8156	0.1707	0.5425	0.5463	0.2403
BAYES	0.9084	0.9484	0.1079	0.7693	0.8025	0.1642	0.7893	0.8337	0.1834

Table 2.5 gives numerical summaries of the relative efficiencies. When the variance is constant over the region of interest, all three methods do a good job, with the medians and means of the relative efficiency above 0.90. As the variance structure becomes complicated, their performances degrade differently. In the Nonconstant Stdev case, the results using three methods are not as good as those in the constant variance case but their performances are still fairly comparable. When the true variance has the logarithm structure, the BAYES model stands out among the three with the largest mean, median and the smallest variance of the relative efficiency. Case-wise speaking, the performance of the BAYES model is much more stable than the two least squares methods as its performance does not decrease as dramatically as the OOLS and GOLS models when the variance structure becomes more complicated. Since in dual response surface problems the variance is assumed to be nonconstant, the performance under the latter two scenarios should be given more emphasis. In the Nonconstant Stdev case, the BAYES model displays comparable modelling efficiencies to the least squares methods though it misspecifies the variance model. In the Nonconstant Logvar case, the two least squares methods perform significantly worse due to variance model misspecification. In conclusion, the BAYES method always produces better or at least equivalent results over the three conditions and hence it should be preferred over the two least squares methods.

2.7 Summary and Conclusion

In literature on dual response surface, much work has been devoted to the improvement of optimization criterion, the optimization algorithms and implementations. However, if a model does not fit the true response surfaces well, the optimization results based on such a model do not have much meaning.

In this paper we propose a Bayesian hierarchical model for the dual response system and we assess its performance by comparing it to two existing frequentist methods. In the real data analysis, the two frequentist methods (OOLS and GOLS) pick two optimal locations far away from each other under the same optimization criterion. In simulation studies, the BAYES method shows robust results under various conditions and performs better or at least equivalent compared to OOLS and GOLS. The consistent performance of the BAYES model excludes ambiguity in model building, parameter estimation and optimal conditions prediction.

In practice, the BAYES model retains the advantages of the dual response approach: it has two separate estimated models to monitor both the mean characteristic and the process variability. Another advantage of the BAYES approach is that it can incorporate prior information from previous studies into the current data analysis, which accords with the sequentiality nature of most response surface problems, e.g. see Myers and Montgomery ([26], pp. 10). Moreover, as the Bayesian methodology is directly based on simulations from probability distributions, it can estimate the predictive distribution of the response at any point within the region of interest. Hence analysts will be given more comprehensive views of the response of interest rather than merely the mean and the variance.

2.8 Appendix

2.8.1 Posterior distributions (conventional informative priors)

Theorem 2.1. Suppose $y_{ij}|\underline{\beta}, \sigma_i^2 \sim N(\underline{x}_i'\underline{\beta}, \sigma_i^2)$ and $\ln \sigma_i^2|\underline{\gamma}, \delta^2 \sim N(\underline{z}_i'\underline{\gamma}, \delta^2)$. Assume the following prior distributions for $\underline{\beta}$, $\underline{\gamma}$, and δ^2:

$$\pi(\underline{\beta}) \sim \text{MVN}_p(\underline{\mu}_\beta, \Sigma_\beta), \quad \pi(\underline{\gamma}) \sim \text{MVN}_q(\underline{\mu}_\gamma, \Sigma_\gamma), \quad \pi(\delta^2) \sim \text{Inverse-Gamma}(\alpha, \tau),$$

where $\underline{\mu}_\beta$, Σ_β, $\underline{\mu}_\gamma$, Σ_γ, α, and τ are the respective mean vectors and covariance matrices of $\underline{\beta}$ and $\underline{\gamma}$ in the prior distributions, and α, τ are the shape and scale parameters of the Inverse-Gamma distribution, respectively. Then the joint posterior distribution of $(\underline{\beta}, \underline{\gamma}, \sigma_1^2, \ldots, \sigma_N^2, \delta^2)$ is:

$$\begin{aligned}
&\pi(\underline{\beta}, \underline{\gamma}, \sigma_1^2, \ldots, \sigma_N^2, \delta^2|Y) \\
&\propto \frac{1}{(\delta^2)^{N/2+\alpha+1}|\Sigma_\beta|^{1/2}|\Sigma_\gamma|^{1/2}\prod_{i=1}^N(\sigma_i^2)^{n_i/2+1}} \exp\left[-\sum_{i=1}^N \frac{(n_i-1)s_i^2 + n_i(\bar{y}_i - \underline{x}_i'\underline{\beta})^2}{2\sigma_i^2}\right] \\
&\times \exp\left[-\frac{1}{2\delta^2}(\underline{d} - Z\underline{\gamma})'(\underline{d} - Z\underline{\gamma}) - \frac{1}{2}(\underline{\beta} - \underline{\mu}_\beta)'\Sigma_\beta^{-1}(\underline{\beta} - \underline{\mu}_\beta) - \frac{1}{2}(\underline{\gamma} - \underline{\mu}_\gamma)'\Sigma_\gamma^{-1}(\underline{\gamma} - \underline{\mu}_\gamma) - \frac{\tau}{\delta^2}\right],
\end{aligned}$$

where $\underline{d} = (d_1, d_2, \ldots, d_N)' = (\ln \sigma_1^2, \ln \sigma_2^2, \ldots, \ln \sigma_N^2)'$.

Proof. Straightforward posterior calculation. □

Theorem 2.2. Suppose we have the same assumptions as in Theorem 2.1. Let $\bar{Y}$ be the vector of sample means at design points, $\bar{Y} = (\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_N)'$, V the variance covariance matrix for $\bar{Y}$, and $\underline{t} = (t_1, t_2, \ldots, t_N)'$. The full conditional posterior distributions are:

$$\begin{aligned}
\pi(\underline{\beta}|\text{others}) &\sim \text{MVN}_p\left((X'V^{-1}X + \Sigma_\beta^{-1})^{-1}(X'V^{-1}\bar{Y} + \Sigma_\beta^{-1}\underline{\mu}_\beta), (X'V^{-1}X + \Sigma_\beta^{-1})^{-1}\right), \\
\pi(\underline{\gamma}|\text{others}) &\sim \text{MVN}_q\left((Z'Z/\delta^2 + \Sigma_\gamma^{-1})^{-1}(Z'\underline{d}/\delta^2 + \Sigma_\gamma^{-1}\underline{\mu}_\gamma), (Z'Z/\delta^2 + \Sigma_\gamma^{-1})^{-1}\right), \\
\pi(\delta^2|\text{others}) &\sim \text{Inverse-Gamma}\left(N/2 + \alpha, \frac{(\underline{d} - Z\underline{\gamma})'(\underline{d} - Z\underline{\gamma}) + 2\tau}{2}\right), \\
\pi(\sigma_i^2|\text{others}) &\propto \frac{1}{(\sigma_i^2)^{n_i/2+1}} \exp\left\{-\frac{1}{2\sigma_i^2}[(n_i-1)s_i^2 + n_i(\bar{y}_i - \underline{x}_i'\underline{\beta})^2] - \frac{1}{2\delta^2}(\ln \sigma_i^2 - \underline{z}_i'\underline{\gamma})^2\right\}, \\
&\quad \text{for} \quad i = 1, 2, \ldots, N.
\end{aligned}$$

Proof. To derive the conditional posterior distribution for certain variable(s), we simply keep terms involving the variable(s) of interest, and treat the other terms from the joint posterior distributions as constants. □

2.8.2 Posterior distributions (noninformative priors)

Corollary 2.1. Suppose $y_{ij}|\underline{\beta}, \sigma_i^2 \sim N(\underline{x}_i'\underline{\beta}, \sigma_i^2)$ and $\ln \sigma_i^2|\underline{\gamma}, \delta^2 \sim N(\underline{z}_i'\underline{\gamma}, \delta^2)$. In the absence of historical information, the following prior distributions are assumed for $\underline{\beta}$, $\underline{\gamma}$, and δ^2:

$$\pi(\underline{\beta}) \propto 1 \quad \pi(\underline{\gamma}) \propto 1 \quad \pi(\delta^2) \propto \frac{1}{\delta^2}\exp(-\frac{\lambda}{\delta^2}),$$

where λ is a small positive number. Then the joint posterior distribution of $(\underline{\beta}, \underline{\gamma}, \sigma_1^2, \ldots, \sigma_N^2, \delta^2)$ is:

$$\pi(\underline{\beta}, \underline{\gamma}, \sigma_1^2, \ldots, \sigma_N^2, \delta^2|Y)$$
$$\propto \frac{1}{(\delta^2)^{N/2+1}\prod_{i=1}^N(\sigma_i^2)^{n_i/2+1}} \exp\left[-\sum_{i=1}^N \frac{(n_i-1)s_i^2 + n_i(\bar{y}_i - \underline{x}_i'\underline{\beta})^2}{2\sigma_i^2} - \frac{(\underline{d}-Z\underline{\gamma})'(\underline{d}-Z\underline{\gamma}) + 2\lambda}{2\delta^2}\right].$$

Corollary 2.2. Suppose we have the same assumptions as in Corollary 2.1. The full conditional posterior distributions are:

$$\begin{aligned}
\pi(\underline{\beta}|\text{others}) &\sim \text{MVN}_p((X'V^{-1}X)^{-1}X'V^{-1}\bar{Y}, (X'V^{-1}X)^{-1}),\\
\pi(\underline{\gamma}|\text{others}) &\sim \text{MVN}_q((Z'Z)^{-1}Z'\underline{d}, (Z'Z)^{-1}\delta^2),\\
\pi(\delta^2|\text{others}) &\sim \text{Inverse-Gamma}\left(\frac{N}{2}, \frac{(\underline{d}-Z\underline{\gamma})'(\underline{d}-Z\underline{\gamma}) + 2\lambda}{2}\right),\\
\pi(\sigma_i^2|\text{others}) &\propto \frac{1}{(\sigma_i^2)^{n_i/2+1}}\exp\left\{-\frac{1}{2\sigma_i^2}[(n_i-1)s_i^2 + n_i(\bar{y}_i - \underline{x}_i'\underline{\beta})^2] - \frac{1}{2\delta^2}(\ln\sigma_i^2 - \underline{z}_i'\underline{\gamma})^2\right\},\\
&\quad \text{for} \quad i = 1, 2, \ldots, N.
\end{aligned}$$

Both Corollaries can be proven easily by using Theorems 2.1 and 2.2.

2.8.3 Propriety of the posterior distributions (noninformative priors)

Theorem 2.3. Let $\Omega_{m\times m}$ be a real symmetric positive definite matrix having eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_m$ with associated orthonormal $(m \times 1)$ eigenvectors $v_1, v_2, \cdots, v_m$. Then

$$\min\{\det(B'\Omega B) : B_{m\times n} \quad \text{real}, B'B = I_n\} = \lambda_1\lambda_2\cdots\lambda_n,$$

where $B = [v_1, v_2, \cdots, v_n]$.

Theorem 2.3 and its proof can be found in Horn and Johnson ([20], pp. 73).

Theorem 2.4. Suppose we have the same assumptions as in Corollary 2.1. Then the posterior distribution of $(\underline{\beta}, \underline{\gamma}, \sigma_1^2, \ldots, \sigma_N^2, \delta^2)$ is proper when $n_i > 1 \quad \text{for} \quad i = 1, 2, \ldots, N$.

Proof. Suppose the mean model matrix X is of full row rank p, then $X'X$ is an $p \times p$ symmetric matrix with corresponding eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_p$ and normalized eigenvectors $c_1, c_2, \cdots, c_p$. With spectral decomposition, $X'X$ can be expressed as

$$X'X = C\Lambda C' \tag{2.2}$$

where $\Lambda = \text{diag}\{\lambda_i\}$, a diagonal matrix with λ_i as the i^{th} element, and C is the orthogonal matrix $C = (c_1, c_2, \cdots, c_p)$. Since $C'C = I_p$, (2.2) can be written as

$$C'X'XC = \Lambda,$$

and then

$$\Lambda^{-\frac{1}{2}}C'X'XC\Lambda^{-\frac{1}{2}} = B'B = I_p, \tag{2.3}$$

where $B = XC\Lambda^{-\frac{1}{2}}$ and $\Lambda^{-\frac{1}{2}} = \text{diag}\{\lambda_i^{-\frac{1}{2}}\}$.

It is known from Corollary 2.1 that the joint posterior distribution of $(\underline{\beta}, \underline{\gamma}, \sigma_1^2, \ldots, \sigma_N^2, \delta^2)$ is:

$$\pi(\underline{\beta}, \underline{\gamma}, \sigma_1^2, \ldots, \sigma_N^2, \delta^2|Y)$$
$$\propto \frac{1}{(\delta^2)^{N/2+1}\prod_{i=1}^N(\sigma_i^2)^{n_i/2+1}} \exp\left[-\sum_{i=1}^N \frac{(n_i-1)s_i^2 + n_i(\bar{y}_i - \underline{x}_i'\underline{\beta})^2}{2\sigma_i^2} - \frac{(\underline{d} - Z\underline{\gamma})'(\underline{d} - Z\underline{\gamma}) + 2\lambda}{2\delta^2}\right].$$

After integrating over $\underline{\gamma}$, δ^2 and $\underline{\beta}$, we have

$$\pi(\sigma_1^2,\ldots,\sigma_N^2|Y) = \frac{\mathcal{C}_1}{|X'D^{-1}X|^{1/2}[\underline{d}'(I-M_z)\underline{d}+2\lambda]^{\frac{N-q}{2}}\prod_{i=1}^N(\sigma_i^2)^{n_i/2+1}} \times \exp\left[-\sum_{i=1}^N\frac{(n_i-1)s_i^2}{2\sigma_i^2}-\frac{1}{2}\overline{\underline{Y}}'(D^{-1}-M_{x,D})\overline{\underline{Y}}\right],$$

where $M_z = Z(Z'Z)^{-1}Z'$, $D=\text{diag}(\sigma_i^2/n_i)$, $M_{x,D}=D^{-1}X(X'D^{-1}X)^{-1}X'D^{-1}$ and $\mathcal{C}_1$ is the normalizing constant in the marginal posterior distribution of $(\sigma_1^2,\ldots,\sigma_N^2)$.

Since both $(I-M_z)$ and $(D^{-1}-M_{x,D})$ are positive semi-definite, $\underline{d}'(I-M_z)\underline{d}$ and $\overline{\underline{Y}}'(D^{-1}-M_{x,D})\overline{\underline{Y}}$ are nonnegative, therefore

$$\pi(\sigma_1^2,\ldots,\sigma_N^2|Y) \le \frac{\mathcal{C}_1\mathcal{C}_2}{|X'D^{-1}X|^{1/2}\prod_{i=1}^N(\sigma_i^2)^{n_i/2+1}}\exp\left[-\sum_{i=1}^N\frac{(n_i-1)s_i^2}{2\sigma_i^2}\right], \tag{2.4}$$

where $\mathcal{C}_2=(2\lambda)^{\frac{N-q}{2}}$. Let

$$\frac{1}{\eta_i}=\frac{(n_i-1)s_i^2}{\sigma_i^2}, \quad \text{for} \quad i=1,2,\cdots,N.$$

Order η's such that

$$\eta_{(1)}\ge\eta_{(2)}\ge\cdots\ge\eta_{(N)} \quad \text{and} \quad \frac{1}{\eta_{(1)}}\le\frac{1}{\eta_{(2)}}\le\cdots\le\frac{1}{\eta_{(N)}}.$$

Then

$$\begin{aligned} X'D^{-1}X &= X'\text{diag}\{\frac{n_i}{\sigma_i^2}\}X = X'\text{diag}\left\{\frac{n_i}{(n_i-1)s_i^2\eta_i}\right\}X \\ &= \left(\text{diag}\left\{\sqrt{\frac{n_i}{(n_i-1)s_i^2}}\right\}X\right)'\text{diag}\left\{\frac{1}{\eta_i}\right\}\left(\text{diag}\left\{\sqrt{\frac{n_i}{(n_i-1)s_i^2}}\right\}X\right). \end{aligned}$$

Suppose the diagonal matrix $\Lambda=\text{diag}\{\lambda_i\}$ consists of the eigenvalues of the symmetric matrix $\left(\text{diag}\left\{\sqrt{\frac{n_i}{(n_i-1)s_i^2}}\right\}X\right)'\left(\text{diag}\left\{\sqrt{\frac{n_i}{(n_i-1)s_i^2}}\right\}X\right)$, and the orthogonal matrix $P=(p_1,p_2,\cdots,p_N)$ contains the corresponding normalized eigenvectors. Take the determinant of $X'D^{-1}X$:

$$\begin{aligned} |X'D^{-1}X| &= \left|\left(\text{diag}\left\{\sqrt{\frac{n_i}{(n_i-1)s_i^2}}\right\}X\right)'\text{diag}\left\{\frac{1}{\eta_i}\right\}\left(\text{diag}\left\{\sqrt{\frac{n_i}{(n_i-1)s_i^2}}\right\}X\right)\right| \\ &= |\Lambda|\cdot|\Lambda^{-1/2}|\cdot|P'|\cdot\left|X_*'\text{diag}\left\{\frac{1}{\eta_i}\right\}X_*\right|\cdot|P|\cdot|\Lambda^{-1/2}| \\ &= |\Lambda|\cdot\left|\Lambda^{-1/2}P'X_*'\text{diag}\left\{\frac{1}{\eta_i}\right\}X_*P\Lambda^{-1/2}\right| = |\Lambda|\cdot\left|U'\text{diag}\left\{\frac{1}{\eta_i}\right\}U\right|, \end{aligned}$$

where $X_* = \text{diag}\left\{\sqrt{\frac{n_i}{(n_i-1)s_i^2}}\right\} X$, and $U = X_* P \Lambda^{-1/2}$. With (2.3), it is known that $U'U = \Lambda^{-1/2} P' X_*' X_* P \Lambda^{-1/2} = I$. Use Theorem 2.3:

$$|X'D^{-1}X|^{1/2} = |\Lambda|^{1/2} \cdot \left| U' \text{diag}\left\{\frac{1}{\eta_i}\right\} U \right| \geq |\Lambda|^{1/2} \cdot \prod_{i=1}^{p} \frac{1}{\eta_{(i)}}.$$

Let $f^*(\sigma_1^2, \cdots, \sigma_N^2)$ denote the right side of (2.4), then

$$\begin{aligned} f^*(\sigma_1^2, \cdots, \sigma_N^2) &\leq \frac{\mathcal{C}_1 \mathcal{C}_2 \mathcal{C}_3}{\left(\prod_{i=1}^{p} \frac{1}{\eta_{(i)}}\right)^{1/2} \left(\prod_{i=1}^{N} \eta_i^{n_i/2+1}\right)} \exp\left(-\sum_{i=1}^{N} \frac{1}{2\eta_i}\right) \\ &= \frac{\mathcal{C}_1 \mathcal{C}_2 \mathcal{C}_3}{\left(\prod_{i=p+1}^{N} \eta_{(i)}\right)^{1/2} \left(\prod_{i=1}^{N} \eta_i^{(n_i+1)/2}\right)} \exp\left(-\sum_{i=1}^{N} \frac{1}{2\eta_i}\right), \end{aligned} \tag{2.5}$$

where $\mathcal{C}_3 = \prod_{i=1}^{N} [(n_i - 1)s_i^2]^{n_i/2+1} / |\Lambda|^{1/2}$.

Break the exponential form $\exp\left(-\sum_{i=1}^{N} \frac{1}{2\eta_i}\right)$ into two parts and write one of them in terms of order statistics $\eta_{(i)}$, then the right hand side of (2.5) becomes:

$$\begin{aligned} & \mathcal{C}_1 \mathcal{C}_2 \mathcal{C}_3 \cdot \frac{1}{\prod_{i=p+1}^{N} \eta_{(i)}^{1/2}} \exp\left(-\sum_{i=1}^{N} \frac{1}{4\eta_{(i)}}\right) \cdot \frac{1}{\prod_{i=1}^{N} \eta_i^{(n_i+1)/2}} \exp\left(-\sum_{i=1}^{N} \frac{1}{4\eta_i}\right) \\ \leq\ & \mathcal{C}_1 \mathcal{C}_2 \mathcal{C}_3 \cdot \frac{1}{\prod_{i=p+1}^{N} \eta_{(i)}^{1/2}} \exp\left(-\sum_{i=p+1}^{N} \frac{1}{4\eta_{(i)}}\right) \cdot \frac{1}{\prod_{i=1}^{N} \eta_i^{(n_i+1)/2}} \exp\left(-\sum_{i=1}^{N} \frac{1}{4\eta_i}\right) \\ \leq\ & (2e^{-1/2})^{N-p} \mathcal{C}_1 \mathcal{C}_2 \mathcal{C}_3 \cdot \frac{1}{\prod_{i=1}^{N} \eta_i^{(n_i+1)/2}} \exp\left(-\sum_{i=1}^{N} \frac{1}{4\eta_i}\right), \end{aligned} \tag{2.6}$$

since $\max\left(\frac{1}{\eta_{(i)}} \exp(-\frac{1}{4\eta_{(i)}})\right) = 4e^{-1}$ for any i.

The right hand side of (2.6) is integrable when all n_i's are greater than 1, therefore the posterior distribution is also integrable and thus proper. □

Chapter 3

Bayesian Hierarchical Modelling on Dual Response Surfaces in Partially Replicated Designs

Abstract

In dual response surface methodology both the mean and variance functions are estimated to monitor an industrial process. Statisticians have discovered that often the variance function follows a lower order model than the mean function (Vining and Schaub [41]), which requires only a portion of the design to be replicated. This paper extends the application of the Bayesian hierarchical model on dual response surfaces in fully replicated experiments (Chen and Ye [6]) to partially replicated designs. Partial data from a well-known experiment are used for illustration and the performance of the Bayesian model is compared with least squares methods by using simulated data under different variance scenarios. Apart from the optimal point selected based on the point estimates of the model parameters, the point with the largest probability of falling inside a certain tolerance region is located to consider the uncertainty of the model parameters.

Key Words: Bayesian hierarchical model; robust parameter design; the Gibbs sampling; genetic algorithm; tolerance region; optimization.

3.1 Introduction

Robust parameter design (RPD) has been successfully used as a cost-effective approach to improve the quality of products and processes since the mid-1980s (Taguchi [36] and Nair [28]). The goal of this methodology is to choose the levels of a certain set of the control factors that optimize a defined process characteristic and reduce the sensitivity of the process to the variations in another set of uncontrollable factors. The primary motivation for the development of RPD is that the quality of a product or a process is characterized not only by the mean but also the variance.

The use of response surface methodology (RSM) for RPD can be traced back to the early 1990s when Vining and Myers ([40]) first propose the dual response surface (DRS) methodology as an alternative approach to Taguchi's crossed-array designs and signal-to-noise ratios. Vining and Myers ([40]) suggest that the process characteristic and its variability form a dual response system, and two separate models are built for the response and its variance. In that paper a full second order design is replicated and two complete second order models are built for the sample mean and the sample standard deviation respectively. The model parameters are estimated with the ordinary least squares method. Vining and Myers ([40]) mention that the generalized least squares (GLS) method should be used to estimate the coefficients in the mean model in order to account for the heterogeneity of variances in dual response problems.

One pitfall of such a modelling approach is that it is possible to yield negative predicted values of the standard deviation, even if the true mean standard deviations are positive throughout the region of interest. If the response at a point with the negative predicted standard deviation happened to be the optimum under certain criterion, it would create difficulty in explaining the process performance at the picked optimal point, and cause confusion among the practitioners. More discussions can be found in Chen and Ye ([6]).

Since Vining and Myers ([40]), a great number of articles have been published regarding the optimization of the dual response system, such as Del Castillo and Montgomery ([12]), Lin and Tu ([24]), Copeland and Nelson ([9]), Ames *et al.* ([2]), Kim and Lin ([23]), and Tang and Xu ([37]). The optimization work in the aforementioned articles focuses mostly on achieving a nominal target value for the mean response and minimizing the variance. Quite often, the performance requirement for a quality characteristic is specified by a tolerance region so that the unavoidable variation in a manufacturing process could be considered. The tolerance region is set by industrial standards or customer requirements. The probability of a process characteristic falling inside the tolerance region is a natural measure for quality and reliability, and thus is referred to as *the quality and reliability measure*.

For the modelling and estimation, Chen and Ye ([6]) propose a Bayesian hierarchical approach for fully replicated second order designs and have shown that the Bayesian method gives better overall performance than two frequentist least squares methods. The comparisons are made under different variance scenarios. This paper extends the work of Chen and Ye ([6]) to partially replicated dual response designs. As in Chen and Ye's article, the estimates of the model parameters are based on posterior inference and a hybrid method that combines of the genetic algorithm and local optimization algorithms is used to find optimal operating conditions based on point estimates. Furthermore, the adaptive genetic algorithm (AGA) ([35]) is adopted to maximize the quality and reliability measure after Bayesian inference.

The paper is organized as follows. Section 3.2 gives an introduction to the genetic algorithm, the hybrid optimization method for DRS, and optimization of the quality and reliability measure. In Section 3.3, the Bayesian hierarchical approach proposed by Chen and Ye ([6]) is sketched and is extended to model the data from partially replicated DRS designs. Its associated computational issues and requirements for the designs are also discussed. After the model development, the proposed model is compared with the least squares methods by using a part of one real data set in Section 3.4 and using simulated data under different variance scenarios in Section 3.5. Section 3.6 contains discussion, and theoretical details of the Bayesian method are placed in the Appendix 3.7.

3.2 Application of the Genetic Algorithm in DRS and Optimization of the Quality and Reliability Measure

3.2.1 The genetic algorithm and hybrid computation methods

The genetic algorithm (GA) is pioneered by Holland ([19]), and applications of GA to problems of mathematical optimizations owe much to De Jong ([10]). GA attempts to mimic the natural evolution of a population by allowing solutions to reproduce, to create new solutions, and to compete for survival in the next iteration. It is done, iteratively, by applying principles of "Darwinian natural selection" to populations of computer representations of the solution domain. After many iterations, the best solution is usually near the global optimum. Instead of the point-to-point search in the traditional methods, GA proceeds from one population to another, and thus it sweeps through the parameter space in many directions simultaneously and thereby reduces the probability of convergence to false optima.

However, the solution from GA does not guarantee the exact global optimum, but a near-global optimum. If a very precise result is required, GA may suffer from excessively slow convergence because of its fundamental requirement of neglecting the local information. If the function to be optimized is expressed in an analytical form, the result from GA can be refined by using gradient-type algorithms. The proposed hybrid method is to first employ GA to find a near-global optimal point, and then the found point serves as the starting point for the local optimization algorithm to identify the exact global optimum. In this paper, the genetic algorithm is designed and developed using Matlab especially for the MSE criterion ([24]) in dual response surface problems. The subsequent local optimization is conducted with SQP methods ([15]) using a Matlab built-in function.

3.2.2 Optimization of the quality and reliability measure

A typical tolerance region for a process characteristic often requires that the response be between the lower bound L and the upper bound U. The objective is to find the operating condition, represented as $\underline{x}_0$, under which the probability $Pr(L \leq y_0 \leq U)$ is maximized over the experimental region. To compute and maximize the quality and reliability measure, we need complete knowledge on the statistical distribution of the characteristic at the location $\underline{x}_0$, that is, essentially the distributions of the model parameters and the variance at that location. However, due to the hierarchical structure in the dual response system, it is very difficult to use the frequentist approach to obtain the distributions of these parameters and the variances, even with the normality assumption on the response and the transformed variance.

As will be seen in Section 3.3, Bayesian analysis is based on inference from the posterior distributions of parameters. With the random deviates of the posterior distributions of the model parameters available, the probability can be well approximated using Monte Carlo simulation for any design point $\underline{x}_0$ in the experimental region using

$$P(\underline{x}_0) \approx \frac{1}{M} \sum_{l=1}^{M} I(L \leq y_{0,l} \leq U), \tag{3.1}$$

where M is the number of iterations in the posterior simulation, $y_{0,l}$ is the response generated conditional on the location $\underline{x}_0$ and the l^{th} simulated set of the model parameters. Here $I(\cdot)$ is an indicator function with 1 corresponding to values inside the tolerance region and 0 to the values outside the region.

For a small number of control factors it is computationally reasonable to grid over the experimental region to compute $P(\underline{x}_0)$ and find the point with the largest P. However, for three or more factors, the grid-over method may suffer from expensive and time-consuming computation. Peterson ([30]) proposes a logistic regression approach as an efficient way to maximize $P(\underline{x}_0)$ in the context of multiple response surface optimization, which can be easily borrowed for DRS optimization. In that approach, $P(\underline{x}_0)$ is calculated for design points over a coarsely dense grid over the experimental region and then a logistic regression model is estimated to obtain an approximate surface for the conformance proportion. With a closed-form model available, a canonical analysis or classical ridge analysis can be done to explore how $P(\underline{x}_0)$ changes as $\underline{x}_0$ moves inside the experimental region.

In the above logistic regression approach, if the model built for the conformance proportion does fit the true surface well, the point chosen by the canonical analysis or ridge analysis could very likely not be the one with the maximum conformance proportion. As pointed out in Peterson's article, if the entire experimental region can not be well fitted with one logistic regression, the Monte Carlo process needs to be re-run with finer grids so that a refined and smaller model can be fit over each sub-region. The whole process could be quite tedious and only verification runs from the Monte Carlo simulation can tell us when the estimated models are good enough for the true surface.

Since GA does not require a closed-form function for optimization, it asks for no estimated models and thus can be used directly. Compared to the functions optimized in [6], the surface of the quality and reliability measure to be maximized here is special in two aspects. First, the computation of the optimization function is much more expensive: the probability at each location is calculated through Monte Carlo simulation, which requires a large sample of the parameters from the posterior distribution and many evaluations of the indicator function in (3.1); while the MSE and the variance functions minimized in [6] are very easy to evaluate. Second, the surface of the quality and reliability is like the sea surface, dynamic but structured, as the probability is calculated based on the observations from the posterior distribution; while the surfaces of the MSE and the variance are static conditional on the point estimates of the model parameters. Due to the speciality of the probability surface, GA is required to converge to the global optimum region in as few generations as possible, and to be robust to the wavy landscape.

The power of GA arises from two genetic operations: crossover and mutation. The significant effect of the crossover rate and the mutation rate upon GA performance has long been acknowledged in GA research ([10]). The higher the crossover rate, the quicker are the new solutions introduced into the population. As the crossover rate increases, solutions can be disrupted faster than selection can exploit them. And large values of the mutation rate can transform GA into a purely random search algorithm. In standard GA, the two rates are specially tailored for the surface to be explored and held constant through the optimization process. Srinivas and Patnaik ([35]) propose the use of adaptive probabilities of crossover and mutation in GA for multimodal function optimization. In the adaptive genetic algorithm (AGA), the crossover rate and the mutation rate are varied depending on the fitness values of the solutions. High fitness solutions are protected with low crossover rate and mutation rate, while solutions with low fitness are totally disrupted. As a result, when a population converges to the global optimum region, the probabilities of crossover and mutation decrease and those high fitness solutions are well preserved. In this paper AGA is used to optimize the quality and reliability measure in Section 3.4.

3.3 Bayesian Hierarchical Modelling in DRS

3.3.1 The Bayesian hierarchical model

As discussed in Chen and Ye ([6]), the Bayesian hierarchical approach is a natural way for modelling the dual response surfaces. Let $\underline{x}$ represent a $k \times 1$ vector of independent control factors, $\mathbf{X}$ an $N \times p$ matrix used in the mean response model and $\mathbf{Z}$ an $N \times q$ matrix in the variance model, where N is the number of distinct design points in the experiment, p and q are the numbers of parameters in the mean and variance models respectively; and $\underline{x}_i'$, $\underline{z}_i'$ are the respective row vectors associated with the matrices $\mathbf{X}$ and $\mathbf{Z}$ for the i^{th} location in the experiment. Denote by $\mathbf{Y} = (y_{11}, y_{12}, \ldots, y_{1n_1}, y_{21}, \ldots, y_{Nn_N})'$ the vector of observations from the experiment, and y_{ij} the j^{th} response observed at the i^{th} design point, where $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, n_i$; and the sample variance and population variance at the i^{th} design point are denoted by s_i^2 and σ_i^2 respectively. Furthermore, $\underline{\beta}$ represents a $p \times 1$ vector of coefficients of the mean response model, $\underline{\gamma}$ a $q \times 1$ vector of coefficients of the variance response model, and δ^2 the hyperparameter in the distribution of σ_i^2's.

The following assumptions on the independence among observations and parameters are made for simplicity, though it may need to be modified in the presence of subjective information.:

- Independence among y_{ij}'s conditional on $\underline{\beta}$ and σ_i^2's:

$$corr(y_{ij}, y_{i'j'}|\underline{\beta}, \sigma_i^2, \sigma_{i'}^2) = 0, \quad \text{for either} \quad i \neq i' \quad \text{or} \quad j \neq j' \quad \text{or both.}$$

- Independence among σ_i^2's conditional on $\underline{\gamma}$ and δ^2:

$$corr(\sigma_i^2, \sigma_{i'}^2|\underline{\gamma}, \delta^2) = 0, \quad \text{for} \quad i \neq i'.$$

- Independence among $\underline{\beta}$, $\underline{\gamma}$, and δ^2 in the prior specification:

$$\pi(\underline{\beta}, \underline{\gamma}, \delta^2) = \pi(\underline{\beta}) \cdot \pi(\underline{\gamma}) \cdot \pi(\delta^2).$$

Then the Bayesian hierarchical model proposed by Chen and Ye ([6]) is developed as follows:

$$\begin{aligned}
&y_{ij}|\underline{\beta}, \sigma_i^2 \sim N(\underline{x}_i'\underline{\beta}, \sigma_i^2), \\
&\ln \sigma_i^2|\underline{\gamma}, \delta^2 \sim N(\underline{z}_i'\underline{\gamma}, \delta^2), \\
&\text{for} \quad i = 1, 2, \ldots, N; \quad j = 1, 2, \ldots, n_i.
\end{aligned}$$

The prior distributions for $\underline{\beta}$, $\underline{\gamma}$, and δ^2 are specified depending on the availability and the form of the subjective information.

Bayesian analysis is based on posterior inference. Estimates of parameters are usually summary statistics of the marginal posterior distributions, such as the posterior mean, median, mode, standard deviation, etc. In simple models, it is often easy to analytically derive the marginal posterior distributions and obtain the summary statistics. However, in more complicated models, especially if the parameters are of multiple dimensions with hierarchical structures, it is often hard or impossible to analytically express the marginal distribution of each parameter. A common strategy is to use simulation techniques, such as the MCMC Gibbs sampling, to generate observations for the marginal posterior distributions and calculate the summary statistics numerically. See Robert and Casella ([33]) for more information.

3.3.2 Extension of the Bayesian hierarchical model to partially replicated designs

If the design is partially replicated, the frequentist least squares methods use only the sample standard deviations of the replicated points in estimating the variance function, although this assumption can be easily relaxed (see Aitkin ([1])). If the unreplicated points need to be employed in estimating the variance model, the residuals between the observed responses and the predicted responses from the estimated mean response model could be used as the response in the variance model. However, the misspecification of the mean model may impact the residuals, and thus further distort the estimation of the variance model. Therefore, researchers often choose to use only the replicated points in estimating the variance model.

Similarly, if m out of N design points have replicates ($m < N$), the Bayesian hierarchical model sets a Log-Normal distribution for σ_i^2 conditional on $\underline{\gamma}$ and δ^2 only when the i^{th} design point has replicates:

$$\ln \sigma_i^2 | \underline{\gamma}, \delta^2 \sim N(\underline{z}_i' \underline{\gamma}, \delta^2) \quad \text{for} \quad i = 1, 2, \ldots, m.$$

For points without replicates, a deterministic function on $\underline{\gamma}$ is assumed:

$$\ln \sigma_i^2 | \underline{\gamma} = \underline{z}_i' \underline{\gamma} \quad \text{for} \quad i = m+1, m+2, \ldots, N.$$

The reason for setting prior distributions on only replicated points is that for locations without replicates we cannot obtain information about the variance directly from a single observation. By using deterministic functions, the variances at the non-replicated points are determined by borrowing information from the information at the replicated points through $\underline{\gamma}$. With this separate variance modelling strategy, we extend the Bayesian hierarchical model from the fully replicated designs to partially replicated designs.

In choosing the prior distributions of $\underline{\beta}$, $\underline{\gamma}$, and δ^2, informative priors are always preferred if prior information with respect to the parameters is available from previous studies or subjective knowledge. The commonly used prior distributions for the above parameters are as follows:

$$\pi(\underline{\beta}) \sim \text{MVN}_p(\underline{\mu}_\beta, \Sigma_\beta) \quad \pi(\underline{\gamma}) \sim \text{MVN}_q(\underline{\mu}_\gamma, \Sigma_\gamma) \quad \pi(\delta^2) \sim \text{Inverse-Gamma}(\alpha, \tau), \qquad (3.2)$$

with $\pi(\delta^2) = \frac{\tau^\alpha}{\Gamma(\alpha)}(\delta^2)^{-(\alpha+1)}\exp(-\tau/\delta^2)$, where $\underline{\mu}_\beta$, Σ_β, and $\underline{\mu}_\gamma$, Σ_γ are the respective mean vectors and covariance matrices of $\underline{\beta}$ and $\underline{\gamma}$ in the prior distributions, and both α, $\tau > 0$, are the shape and scale parameters of the Inverse-Gamma distribution, respectively. For a dual response surface problem without any subjective information on parameters, the following independent prior distributions can be used:

$$\pi(\underline{\beta}) \propto 1 \quad \pi(\underline{\gamma}) \propto 1 \quad \pi(\delta^2) \propto \frac{1}{\delta^2}\exp(-\frac{\lambda}{\delta^2}), \tag{3.3}$$

where λ is a small positive number added to prevent the posterior distribution from being improper. It is expected that adding a small positive number should exert no influence on the parameters estimation. See Chen and Ye ([6]) for a detailed discussion on the effects of λ. If improper priors are adopted, the propriety of the posterior distribution needs to be checked.

The difference in the variance modelling leads to difference in the associated computational and theoretical issues. The posterior simulation is processed with the MCMC Gibbs sampling technique as the marginal posterior distributions of the parameters cannot be analytically expressed due to the complexity of dual response problems. To apply the Gibbs sampling technique, the joint posterior distribution and full conditional posterior distributions for using the conventional proper priors in (3.2) and the improper priors in (3.3) are derived and proved in Appendices 3.7.1 and 3.7.2 respectively.

The propriety of the posterior distribution by using the improper priors in (3.3) are discussed in Appendix 3.7.3. Moreover, as will be seen in the proof of Theorem 3.3, the number of replicated design points needs to be greater than the dimension of $\underline{\gamma}$ (i.e. $m \geq q+1$) so that δ^2 can be integrated as an Inverse-Gamma variate and the posterior distribution is proper. The reason for requiring $m \geq q + 1$ is that, apart from prior information, only the replicated points essentially provide information to estimate the parameters in the variance model. There are total $q + 1$ parameters ($\underline{\gamma}$ and δ^2) in the variance model, so we need at least $q + 1$ supportive points.

3.4 An Example

The data set comes from an experiment involving printing ink and it has been widely applied in dual response surface analysis for illustration, e.g. in Vining and Myers ([40]), Ames *et al.* ([2]), and Copeland and Nelson ([9]). The purpose of the experiment was to study the effects of speed (x_1), pressure (x_2), and distance (x_3), upon a printing machine's ability to apply coloring inks upon package labels. The original experiment used a 3^3 complete factorial design with three runs at each design point for a total of 81 runs. All previous analysis try to minimize the process variability around a target value of 500 for the response. The same objective is considered in this paper. Since the experimental region is cuboidal with each factor taking values at $-1, 0, 1$, optimal conditions are searched for throughout the cuboidal region.

In this section, the proposed Bayesian hierarchical model (henceforth referred as BAYES) is compared with two frequentist methods by using part of the printing ink data, due to the unavailability of data sets from partially replicated dual response designs in the literature. The terms contained in the models are assumed to be known in advance: the mean model is of complete second order and the variance model contains only the linear terms and the intercept. The coefficients in the two models, $\underline{\beta}$ and $\underline{\gamma}$, are estimated with their respective posterior medians. Posterior medians are preferred to the posterior means for their robustness to outliers. The optimization results from the Bayesian model are compared with the two frequentist least squares methods discussed in Vining and Myers ([40]): one is to model the mean and standard deviation with linear regressions and estimate both sets of parameters with ordinary least squares (OOLS); and the other is similar to OOLS except that the mean response model coefficients are estimated with generalized least squares (GOLS).

To compare the optimization result from each model, optimal operating conditions are found by using two criteria in the literature. The first one, proposed by Vining and Myers ([40]), minimizes the variability subject to this constraint that the mean response hit at the target value (henceforth referred to as "target is the best" criterion). The second one, proposed by Lin and Tu ([24]), minimizes the sum of deviance around the target value and the variability. This measure is defined as $(\text{mean} - \text{target})^2 + \text{variance}$, which is very similar to the definition of the mean squared error (MSE criterion).

In Vining and Schaub ([41]), it is concluded that replicating a resolution *III* fraction of the factorial portion of a central composite design is a particularly effective experimental strategy in terms of total design runs and the performance in model estimation. Though the printing ink data set does not exactly come from a central composite design, we still think it is reasonable to replicate the factorial portion of the design. Instead of replicating only the resolution *III* fraction, we replicate the whole factorial portion. This is because a linear variance model in the Bayesian hierarchical approach contains five parameters: the intercept, three linear terms and one variance term. According to the requirement for the design discussed in Section 3.3.2, we at least need five replicated points for the variance model estimation, while a resolution *III* fraction can only provide four supportive points. Therefore, the part of the data set to be analyzed includes all three observations at the eight factorial points, and one observation randomly chosen from the three at the other design points.

It should be noted that although the combination of a replicated factorial design and a linear variance model is adopted in this example, there are many other sensible choices. For the design part, we could choose to replicate the axial portion, or the axial portion plus the center run, or the factorial portion plus the center, or a resolution *III* fraction of the factorial portion plus the center. And for the variance model, we can have a model with all linear and interaction terms, or a model with all linear and quadratic terms, or a model with some, not all, linear, interaction and quadratic terms. The choice for the variance model completely depends on subjective knowledge or historical information on the form of the variance function. In choosing the proper design, we first need the design to be sufficient for the model estimation, and then we consider about the total number of design runs and how the design points should be spread over the experimental region to gain as much information as possible.

Table 3.1 shows the estimation and optimization results from the three models. $\underline{\hat{\beta}}$ and $\underline{\hat{\gamma}}$ are the estimates of the coefficients in the mean and variance model respectively. Values are ordered as the intercept term first and then three linear terms followed by three quadratic

Table 3.1: Optimization results comparison for the printing ink data

Models	OOLS	GOLS	BAYES
$\underline{\hat{\beta}}$	(339.9,176.9,107.1, 123.3,35.6,-33.3,-39.1, 61.0,73.5,38.5)	(357.1,171.0,99.7, 137.1,11.2,-43.3,-24.9, 10.3,72.0,55.7)	(344.8,172.7,114.1, 107.7,17.9,-36.3,-35.4, 13.8,59.3,54.2)
$\underline{\hat{\gamma}}$	(57.1,-0.3,5.2,30.4)	same as OOLS	(6.7,0.2,0.5,1.3)
$\underline{x}_a$	(1.000,1.000,-0.712)	(1.000,0.542,-0.347)	(1.000,0.314,-0.365)
$\hat{\sigma}_a^2 = \widehat{MSE}_a$	1628.9	2409.6	745.09
$\underline{x}_b$	(1.000,1.000,-0.726)	(1.000,0.528,-0.368)	(1.000,0.307,-0.374)
$\hat{\mu}_b$	495.83	494.26	497.70
$\hat{\sigma}_b^2$	1593.6	2342.5	734.39
$\widehat{MSE}_b$	1611.0	2375.5	739.70

terms and three interaction terms, if they are contained in their respective models. $\underline{x}_a$ denotes the optimal location found under the "target is the best" criterion. $\hat{\sigma}_a^2$ and $\widehat{MSE}_a$ are respectively the estimated mean, variance and MSE at the location $\underline{x}_a$. Similarly, notations $\hat{\mu}_b$, $\hat{\sigma}_b^2$ and $\widehat{MSE}_b$ are for the results found under the MSE criterion.

It can be seen that although the OOLS and GOLS models share the estimates of $\underline{\gamma}$, the estimated $\underline{\beta}$ values are quite different, especially in the coefficient estimate of the interaction term x_1x_3 (61.0 in OOLS vs. 10.3 in GOLS). As a result, the optimal location predicted using the two models are not close to each other, neither are the predicted means, variances, and MSEs. In terms of the predicted variances and MSEs at the found locations, the Bayes model dominates with values of 745.09, 734.39, and 739.70 under the "target is the best" and the MSE criterion, respectively.

However, none of the above arguments could be used as reasons to prefer one model to the others. All predicted variances and MSEs depend upon the estimated models. The settings located by each model are "optimal" only if the models used to generate these predictions approximate the nature of the dual response surfaces well, which we have no clue for a real data set. Therefore, we proceed with simulation studies in Section 3.5 to compare their performances under different scenarios.

As mentioned in Section 3.2, if the quality requirement is specified in terms of a tolerance region, we need to know the distribution of the characteristic to calculate the probability that the response falls inside the tolerance region. For the illustration purpose, we suppose the tolerance region for the response in the printing ink data is $(450, 550)$. The determination of the tolerance region is not a statistical problem. It is usually specified by customers or experts.

In the frequentist approach, the quality and reliability measure at the location $\mathbf{x}_0$ is often calculated as $Pr(L \leq y_0 \leq U|\mathbf{x}_0, \underline{\hat{\beta}}, \underline{\hat{\gamma}})$ once $\underline{\hat{\beta}}$ and $\underline{\hat{\gamma}}$ are obtained. This tends to over-estimate the probability as it does not account for the uncertainty of $\underline{\beta}$ and $\underline{\gamma}$. The Bayesian approach computes the probability as $Pr(L \leq y_0 \leq U|\mathbf{x}_0, \text{data})$ to account for the uncertainty of the model parameters as well as the information from the data set. The computation of the probability is based on the posterior distributions of $\underline{\beta}$, $\underline{\gamma}$ and δ^2. The probability at $\mathbf{x}_0$ is calculated as follows. Suppose M sets of $\underline{\beta}$, $\underline{\gamma}$ and δ^2 from their respective posterior distributions are available. For the l^{th} set of parameters, the variance $\sigma^2_{l,0}$ at $\mathbf{x}_0$ is simulated by:

$$\ln \sigma^2_{l,0} \sim N(\underline{z}_i' \underline{\gamma}_l, \delta_l^2),$$

and then the response $y_{l,0}$ is simulated by:

$$y_{l,0} \sim N(\underline{x}_i' \underline{\beta}_l, \sigma^2_{l,0}).$$

After that, the quality and reliability measure can be computed as

$$Pr(L \leq y_0 \leq U|\mathbf{x}_0, \text{data}) \approx \frac{1}{M} \sum_{l=1}^{M} I(L \leq y_{0,l} \leq U|\mathbf{x}_0, \text{data}).$$

AGA is adopted to find the location with the maximum quality and reliability measure. After finding the best point using AGA, the responses at that point are re-simulated and the probability value is re-calibrated. This is because the quality and reliability measure is a stochastic system and thus the upward bias in the maximum value should be expected. Table 3.2 displays the optimization results of the probability from three difference chains. It lists the best location $\mathbf{x}_0$, the probability at $\mathbf{x}_0$ obtained in the optimization process P_{sim}, and the probability obtained through re-calibration P_{re}.

Table 3.2: Using AGA to optimize the quality and reliability measure

Chain	**Location** (x_1, x_2, x_3)	$\mathbf{P}_{\text{sim}}$	$\mathbf{P}_{\text{re}}$
1	(0.9912, 0.0310, -0.2643)	0.6079	0.6075
2	(1.0000, 0.0323, -0.3025)	0.6061	0.6061
3	(0.9917, 0.0348, -0.2575)	0.6071	0.6069

Table 3.2 shows that the results from three chains of AGA are highly consistent: the best locations are very close to each other and the quality and reliability measure values are very similar, which indicates that AGA does an excellent job in converging to the global optimum region. Compared to the standard GA, the number of generations needed before convergence is reduced by approximately 10 times and the global optimum accuracy is also improved by at least 1 decimal point in terms of the coordinates of the found optima.

3.5 Simulation Studies

Chen and Ye ([6]) study the complete replicated case when both the mean and the variance models are of full second order. In this paper, we look into three combinations: a linear variance model with a replicated factorial design; a linear plus interaction variance model with a replicated factorial design; and a linear variance model with a replicated axial design. For each combination, the basic design is 3^3 cubic as in the printing ink data, and the mean response model is of complete second order. As in Chen and Ye ([6]), we simulate data under three different variance scenarios:

- **Constant variance**:

$$y_{ij} \sim N(\underline{x}_i'\underline{\beta}, 1), \quad \text{for} \quad i = 1, 2, \ldots, N, \quad j = 1, 2, \ldots, n_j;$$

- Nonconstant variance where the standard deviation follows a normal distribution (referred to as **Stdev**):

  $$y_{ij} \sim N(\underline{x}_i'\underline{\beta}, \sigma_i^2), \quad \sigma_i \sim N(\underline{z}_i'\underline{\gamma}, \eta^2), \quad \text{for} \quad i = 1, 2, \ldots, N, \quad j = 1, 2, \ldots, n_j.$$

  In generating the standard deviation, the minimum mean standard deviation over the experimental region is set to be much larger than η, so that the simulated standard deviation has little likelihood to be negative;

- Nonconstant variance where the variance follows a log-normal distribution (referred to as **Logvar**):

  $$y_{ij} \sim N(\underline{x}_i'\underline{\beta}, \sigma_i^2), \quad \sigma_i^2 \sim \text{Log-Normal}(\underline{z}_i'\underline{\gamma}, \eta^2), \quad \text{for} \quad i = 1, 2, \ldots, N, \quad j = 1, 2, \ldots, n_j.$$

For each scenario, $\underline{x}_i'$ contains the intercept, linear terms, quadratic terms and interactions in order, and $\underline{z}_i'$ contains the intercept, linear terms, and interactions (if they exist in the model). For simplicity, we would assume that we know in advance what terms would be included in both the mean and variance models. That is, all three methods have no problems of under- or over-parameterizations. For each combination of the design and model, 2500 data sets are generated. The simulation size is determined such that the largest standard error of the mean efficiency is less than 0.005. The optimal operating conditions are searched for over the cuboidal region ($-1 < x_1, x_2, x_3 < 1$) under the MSE criterion.

The relative efficiency, defined as

$$\text{relative efficiency} = \frac{\min MSE_{\text{true}}}{MSE_{\text{point proposed}}},$$

is used as a single-number measure of performance to see how well a model does in choosing "good points." If the operating condition recommended by a model performs closely to the true optimum, i.e., the relative efficiency is very near to 1, then the model is viewed as good. If the relative efficiency of a model is very small, that means the model picks a point far from optimum. The relative efficiency measure is also used in this paper to compare the overall performance of the three methods. Tables 3.3 - 3.5 list the mean, median, and standard deviation of the relative efficiencies for each variance scenario under three different combinations of designs and models.

Apart from the relative efficiency over a large number of data sets, we think that the modelling performance for each data set is also important. Therefore we propose the ratio of the relative efficiencies as a measure to compare how good the three methods are in modelling the same data set. We use the relative efficiency of the BAYES model as the baseline and compare it with those of the other two, with "OOLS/BAYES" being the ratio of the relative efficiency of OOLS against BAYES, and "GOLS/BAYES" the ratio of GOLS against BAYES. If the ratio is larger than 1, it means that OOLS or GOLS is better than BAYES for that specific data set, and vice versa. Figures 3.1 - 3.3 display the boxplots for the logarithm of the ratios of the relative efficiencies. The logarithm of the ratios are plotted so that ratios below "1" can stretch down in the same scale as those above "1" stretch up. When the logarithm of the ratios takes values above 0, then OOLS or GOLS are better, and when below 0, the BAYES method is better. The box itself contains the middle half of the logarithm of the efficiency ratio. The upper hinge of the boxplot indicates the 75^{th} percentile, and the lower hinge indicates the 25^{th} percentile. The ends of the whiskers indicate the minimum and the maximum, unless outliers are present in which case the whiskers extend to a maximum of the 1.5 times the inter-quartile range.

Table 3.3: The relative efficiencies of the three methods in the case of replicated factorial designs and linear variance models. Simulation size is 2500 for each scenario.

Models	**Constant**			**Stdev**			**Logvar**		
	Mean	Median	Stdev	Mean	Median	Stdev	Mean	Median	Stdev
OOLS	0.8532	0.9053	0.1516	0.7532	0.8027	0.1847	0.7073	0.7458	0.2012
GOLS	0.8557	0.9125	0.1533	0.8105	0.8566	0.1516	0.7980	0.8359	0.1554
BAYES	0.8569	0.9121	0.1512	0.8112	0.8688	0.1707	0.8145	0.8494	0.1549

In Table 3.3, where the factorial portion is replicated and the variance models contains only linear terms, all three methods do a good job when the true variance is constant over the region of interest. When the variance is of the Stdev case, the OOLS method is a little worse than the other two; and when the variance is of the Logvar case, the BAYES method is much better than OOLS and slightly better than GOLS. Figure 3.1 compares the performance of

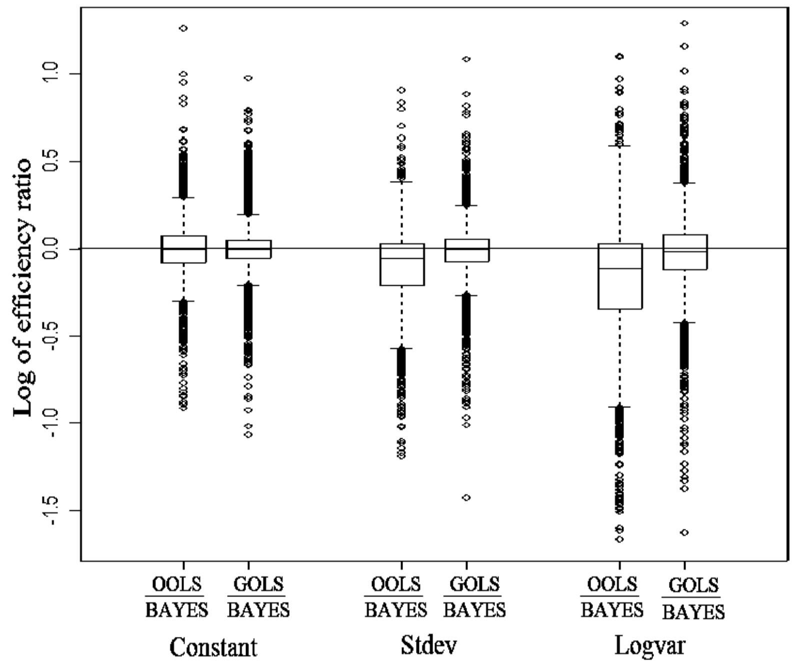

Figure 3.1: Boxplots for the logarithm of the relative efficiency ratios when the factorial portion is replicated and the variance model contains linear terms.

the three methods conditional on the same data. For the ratio GOLS/BAYES when the variance is Constant or Stdev, the bodies of the box plots squeeze around zero, which means that the points between the first quartile and the third quartile are very close to zero, and thus the performances of each pair of methods are similar to each other under the three situations. For the ratio OOLS/BAYES in the Stdev case, the body of the box plot is a little spread below zero, indicating that BAYES is a little better than GOLS, and this trend is even more obvious in for OOLS/BAYES in the Logvar case.

Table 3.4 and Figure 3.2 are for the combination where the design is factorial replicated and the variance model contains the linear and interaction terms. The results show that when the variance is Stdev, the BAYES model is a little worse than GOLS, but still better than OOLS; while when the variance is Logvar, the BAYES model dominates in terms of the mean and median of the relative efficiencies. As seen from Figure 3.2, in the Stdev case, the boxplot for the ratio between GOLS and BAYES moves a little upward from the line zero, indicating the BAYES method is not as good as GOLS. This is probably because

Table 3.4: The relative efficiencies of the three methods in the case of replicated factorial designs and linear plus interaction variance models. Simulation size is 2500 for each scenario.

Models	**Constant**			**Stdev**			**Logvar**		
	Mean	Median	Stdev	Mean	Median	Stdev	Mean	Median	Stdev
OOLS	0.8471	0.9035	0.1580	0.6675	0.6468	0.2185	0.4040	0.3894	0.2201
GOLS	0.8427	0.8995	0.1602	0.7567	0.8119	0.2071	0.4866	0.5121	0.1980
BAYES	0.8478	0.9066	0.1573	0.7305	0.7530	0.1562	0.7811	0.8513	0.2035

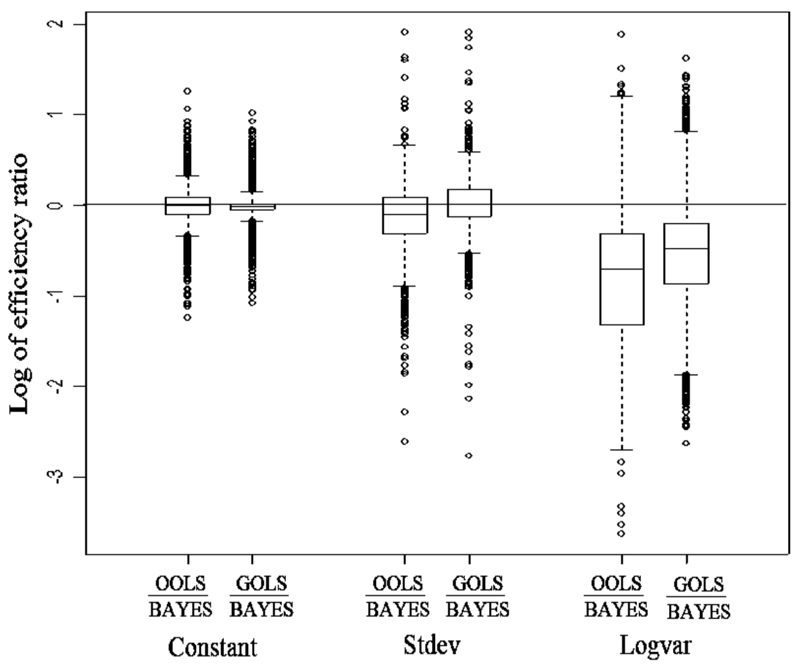

Figure 3.2: Boxplots for the logarithm of the relative efficiency ratios when the factorial portion is replicated and the variance model contains linear and interaction terms.

in this situation, the BAYES method not only has model misspecification, but also the factorial replicated design is saturated for the variance model which contains the linear and the interaction terms. The boxplots for the ratios "OOLS/BAYES" and "GOLS/BAYES" in the Logvar case move downward far away from the zero line, which indicates that under those two conditions, the BAYES model performs much better than the two least squares models.

Table 3.5: The relative efficiencies of the three methods in the case of replicated axial designs and linear variance models. Simulation size is 2500 for each scenario.

Models	**Constant**			**Stdev**			**Logvar**		
	Mean	Median	Stdev	Mean	Median	Stdev	Mean	Median	Stdev
OOLS	0.8285	0.8911	0.1752	0.6442	0.6660	0.2321	0.5434	0.5531	0.2459
GOLS	0.7878	0.8520	0.2066	0.6622	0.7085	0.2320	0.6044	0.6529	0.2470
BAYES	0.8019	0.8598	0.1906	0.7693	0.8209	0.1812	0.7154	0.7691	0.2080

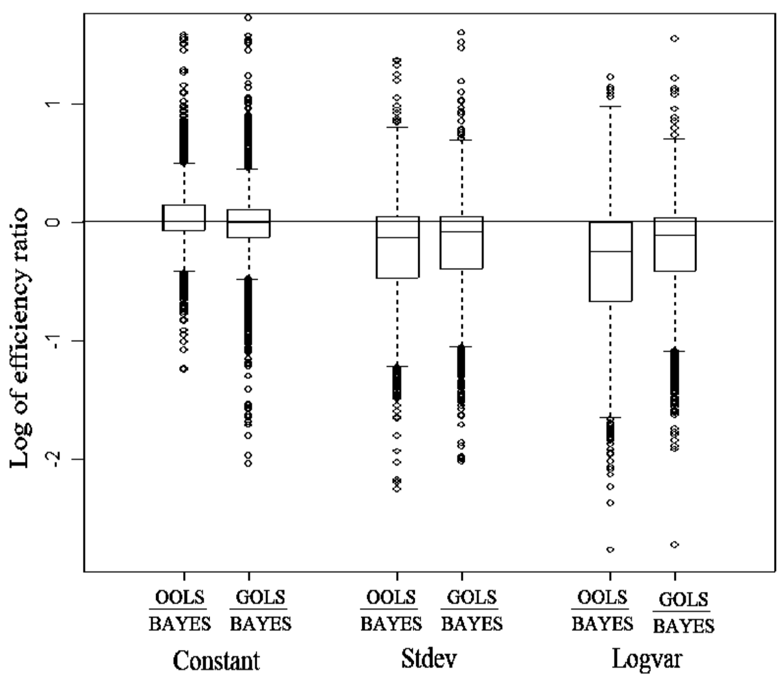

Figure 3.3: Boxplots for the logarithm of the relative efficiency ratios when the axial portion is replicated and the variance model contains linear terms.

Table 3.5 and Figure 3.3 are for the combination where the design is axial replicated and the variance model contains the linear terms only. The performance of the three methods is a little strange in this case: when the variance is constant, the median and mean of the GOLS relative efficiencies are smaller than those of OOLS; and when the variance is Stdev, the relative efficiencies of OOLS and GOLS are significantly lower than the BAYES method, although they have no model misspecification while the BAYES does have. A possible explanation for this phenomenon is that only the axial points are replicated in the design, and those axial points are not spread out to provide OOLS and GOLS enough information on the variance model over the experimental region. In Vining and Schaub ([41]), the axial replicated design strategy is also not recommended based on their simulation studies.

Readers may have noticed that the relative efficiencies of BAYES in the Stdev case are even higher than those in the Logvar case. It may appear uncommon since the BAYES method has model misspecification in the Stdev case while not in the Logvar case. However, it is unfair to compare the performance across different variance scenarios, as the true dual response surfaces are different and it also results in different degrees of the difficulty to approximate the true response.

We can also make comparisons across the two combinations, the factorial replicated design with the linear variance model and the axial replicated design with the linear variance model since the nature of the true dual responses are the same. It is observed that all three methods perform much worse in the axial replicated design than in the factorial replicated design. One reason is that the axial replicated design does not have as many support points for the variance model as in the factorial design, and the other reason is that the replicated axial points are not spread out enough to cover the region of interest.

3.6 Summary and Conclusion

This paper extends the Bayesian hierarchical model for dual response surface from full replicated designs to partially replicated designs. By doing this, the Bayesian hierarchical model can be a complete systematic modelling approach for all types of design in dual response surface.

When the basic design is cubic, a comprehensive comparison between the Bayesian and two least squares methods is carried out under different design strategies, different nature of dual response surfaces, and different variance scenarios. By assessing the overall modelling efficiency and the efficiency conditional on the same data set, the Bayesian method is found to be superior than the two least squares methods in most nonconstant variance cases: in the Stdev situations, the Bayesian method is fairly comparable to the other two; and in the Logvar situations, it is significantly better than the least square methods. Especially when the information from the design is not good enough for the variance modelling (in the axial replicated designs), the Bayesian model still does a better job. The consistently good performance shows that the Bayesian method is more robust, not only to different variance surfaces but also to different designs.

This paper also explores another advantage of the Bayesian approach: the posterior distributions of the model parameters can be used to calculate and optimize the probability that a characteristic falls into a certain tolerance region. AGA is adopted to search the surface of the probability as the surface is stochastic and computationally expensive. It is shown that the application of AGA is straightforward, avoiding ambiguity in establishing models to approximate the true surface. By making the key operation parameters adaptive to fitness values, AGA protects high fitness solutions from being changed, and thus is more efficient in converging to the global optimum region than the conventional GA.

3.7 Appendix

3.7.1 Posterior distributions (conventional informative priors)

Theorem 3.1. Suppose $y_{ij}|\underline{\beta}, \sigma_i^2 \sim N(\underline{x}_i'\underline{\beta}, \sigma_i^2)$ for $i = 1, 2, 3, \ldots, N$, $j = 1, 2, \ldots, n_i$, and $\ln \sigma_i^2|\underline{\gamma}, \delta^2 \sim N(\underline{z}_i'\underline{\gamma}, \delta^2)$ for m replicated design points, where $i = 1, 2, 3, \ldots, m$. The following prior distributions are assumed for $\underline{\beta}$, $\underline{\gamma}$, and δ^2:

$$\pi(\underline{\beta}) \sim \text{MVN}_p(\underline{\mu}_\beta, \Sigma_\beta), \quad \pi(\underline{\gamma}) \sim \text{MVN}_q(\underline{\mu}_\gamma, \Sigma_\gamma), \quad \pi(\delta^2) \sim \text{Inverse-Gamma}(\alpha, \tau),$$

where $\underline{\mu}_\beta$, Σ_β, $\underline{\mu}_\gamma$, Σ_γ, α, and τ are the respective mean vectors and covariance matrices of $\underline{\beta}$ and $\underline{\gamma}$ in the prior distributions, and α, τ are the shape and scale parameters of the Inverse-Gamma distribution, respectively. Then the joint posterior distribution of $(\underline{\beta}, \underline{\gamma}, \sigma_1^2, \ldots, \sigma_N^2, \delta^2)$ is:

$$
\begin{aligned}
&\pi(\underline{\beta}, \underline{\gamma}, \sigma_1^2, \ldots, \sigma_N^2, \delta^2 | Y) \\
\propto &\frac{1}{(\delta^2)^{m/2+\alpha+1} |\Sigma_\beta|^{1/2} |\Sigma_\gamma|^{1/2} \prod_{i=1}^m (\sigma_i^2)^{n_i/2+1} \prod_{i=m+1}^N (\sigma_i^2)^{1/2}} \\
\times &\exp\left[-\sum_{i=1}^m \frac{(n_i - 1)s_i^2 + n_i(\bar{y}_i - \underline{x}_i' \underline{\beta})^2}{2\sigma_i^2} - \sum_{i=m+1}^N \frac{(y_i - \underline{x}_i' \underline{\beta})^2}{2\sigma_i^2} \right] \\
\times &\exp\left[-\frac{1}{2\delta^2}(\underline{d} - Z\underline{\gamma})'(\underline{d} - Z\underline{\gamma}) - \frac{1}{2}(\underline{\beta} - \underline{\mu}_\beta)' \Sigma_\beta^{-1} (\underline{\beta} - \underline{\mu}_\beta) - \frac{1}{2}(\underline{\gamma} - \underline{\mu}_\gamma)' \Sigma_\gamma^{-1} (\underline{\gamma} - \underline{\mu}_\gamma) - \frac{\tau}{\delta^2} \right],
\end{aligned}
$$

where $\underline{d} = (d_1, d_2, \ldots, d_m)' = (\ln \sigma_1^2, \ln \sigma_2^2, \ldots, \ln \sigma_m^2)'$, and $\ln \sigma_i^2 = \underline{z}_i' \underline{\gamma}$ for $i = m+1, \ldots, N$.

Proof. The joint posterior distribution is proportional to the product of the probability density of the data Y and all the prior distributions with independence among observations and parameters assumed.

The distribution of data Y conditional on $\underline{\beta}$ and $(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$ is:

$$
\begin{aligned}
f(Y | \underline{\beta}, \sigma_1^2, \ldots, \sigma_N^2) = &\prod_{i=1}^m \frac{1}{(2\pi\sigma_i^2)^{n_i/2}} \exp\left[-\frac{\sum_{j=1}^{n_i} (y_{ij} - \underline{x}_i' \underline{\beta})^2}{2\sigma_i^2} \right] \\
\times &\prod_{i=m+1}^N \frac{1}{(2\pi\sigma_i^2)^{1/2}} \exp\left[-\frac{(y_i - \underline{x}_i' \underline{\beta})^2}{2\sigma_i^2} \right].
\end{aligned}
$$

The probability density function of σ_i^2 $(i = 1, 2, \ldots, m)$ given $\underline{\gamma}$ and δ^2 is:

$$
f(\sigma_i^2 | \underline{\gamma}, \delta^2) = \frac{1}{\sigma_i^2 \sqrt{2\pi\delta^2}} \exp\left[-\frac{1}{2\delta^2} (\ln \sigma_i^2 - \underline{z}_i' \underline{\gamma})^2 \right].
$$

The product of the above conditional distributions and the assumed prior distributions on $\underline{\beta}$, $\underline{\gamma}$, and δ^2 yield the joint posterior distribution of $(\underline{\beta}, \underline{\gamma}, \sigma_1^2, \ldots, \sigma_N^2, \delta^2)$. □

Theorem 3.2. Suppose we have the same assumptions as in Theorem 3.1. Let $\bar{Y}$ be the vector of sample means at design points, $\bar{Y} = (\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_N)'$, V the variance covariance matrix for $\bar{Y}$, and $\underline{t} = (t_1, t_2, \ldots, t_N)'$. The full conditional posterior distributions are:

$$\begin{aligned}
\pi(\underline{\beta}|\text{others}) &\sim \text{MVN}_p\left((X'V^{-1}X + \Sigma_\beta^{-1})^{-1}(X'V^{-1}\bar{Y} + \Sigma_\beta^{-1}\underline{\mu}_\beta), (X'V^{-1}X + \Sigma_\beta^{-1})^{-1}\right), \\
\pi(\underline{\gamma}|\text{others}) &\sim \text{MVN}_q\left((Z'Z/\delta^2 + \Sigma_\gamma^{-1})^{-1}(Z'\underline{d}/\delta^2 + \Sigma_\gamma^{-1}\underline{\mu}_\gamma), (Z'Z/\delta^2 + \Sigma_\gamma^{-1})^{-1}\right), \\
\pi(\delta^2|\text{others}) &\sim \text{Inverse-Gamma}\left(m/2 + \alpha, \frac{(\underline{d} - Z\underline{\gamma})'(\underline{d} - Z\underline{\gamma}) + 2\tau}{2}\right), \\
\pi(\sigma_i^2|\text{others}) &\propto \frac{1}{(\sigma_i^2)^{n_i/2+1}} \exp\left\{-\frac{1}{2\sigma_i^2}[(n_i - 1)s_i^2 + n_i(\bar{y}_i - \underline{x}_i'\underline{\beta})^2] - \frac{1}{2\delta^2}(\ln\sigma_i^2 - \underline{z}_i'\underline{\gamma})^2\right\}, \\
&\quad \text{for} \quad i = 1, 2, \ldots, m.
\end{aligned}$$

Proof. To derive the conditional posterior distribution for certain variable(s), we simply keep terms involving the variable(s) of interest, and remove the other terms from the joint posterior distributions. □

3.7.2 Posterior distributions (noninformative priors)

Corollary 3.1. Suppose $y_{ij}|\underline{\beta}, \sigma_i^2 \sim N(\underline{x}_i'\underline{\beta}, \sigma_i^2)$ for $i = 1, 2, 3, \ldots, N$, $j = 1, 2, \ldots, n_i$, and $\ln\sigma_i^2|\underline{\gamma}, \delta^2 \sim N(\underline{z}_i'\underline{\gamma}, \delta^2)$ for m replicated designs points, where $i = 1, 2, 3, \ldots, m$. In the absence of subjective information on the parameters $\underline{\beta}$, $\underline{\gamma}$, and δ^2, the following prior distributions are assumed:

$$\pi(\underline{\beta}) \propto 1 \quad \pi(\underline{\gamma}) \propto 1 \quad \pi(\delta^2) \propto \frac{1}{\delta^2}\exp(-\frac{\lambda}{\delta^2}),$$

where λ is a small positive number. Then the joint posterior distribution of $(\underline{\beta}, \underline{\gamma}, \sigma_1^2, \ldots, \sigma_N^2, \delta^2)$ is:

$$\begin{aligned}
&\pi(\underline{\beta}, \underline{\gamma}, \sigma_1^2, \ldots, \sigma_N^2, \delta^2|Y) \\
\propto &\frac{1}{(\delta^2)^{m/2+1}\prod_{i=1}^m(\sigma_i^2)^{n_i/2+1}\prod_{i=m+1}^N(\sigma_i^2)^{1/2}} \exp\left[-\frac{(\underline{d} - Z\underline{\gamma})'(\underline{d} - Z\underline{\gamma}) + 2\lambda}{2\delta^2}\right] \\
\times &\exp\left[-\sum_{i=1}^m \frac{(n_i - 1)s_i^2 + n_i(\bar{y}_i - \underline{x}_i'\underline{\beta})^2}{2\sigma_i^2} - \sum_{i=m+1}^N \frac{(y_i - \underline{x}_i'\underline{\beta})^2}{2\sigma_i^2}\right].
\end{aligned}$$

Corollary 3.2. Suppose we have the same assumptions as in Corollary 3.1. The full conditional posterior distributions are:

$$
\begin{aligned}
\pi(\underline{\beta}|\text{others}) &\sim \text{MVN}_p((X'V^{-1}X)^{-1}X'V^{-1}\bar{Y}, (X'V^{-1}X)^{-1}),\\
\pi(\underline{\gamma}|\text{others}) &\sim \text{MVN}_q((Z'Z)^{-1}Z'\underline{d}, (Z'Z)^{-1}\delta^2),\\
\pi(\delta^2|\text{others}) &\sim \text{Inverse-Gamma}\left(\frac{m}{2}, \frac{(\underline{d}-Z\underline{\gamma})'(\underline{d}-Z\underline{\gamma})+2\lambda}{2}\right),\\
\pi(\sigma_i^2|\text{others}) &\propto \frac{1}{(\sigma_i^2)^{n_i/2+1}}\exp\left\{-\frac{1}{2\sigma_i^2}[(n_i-1)s_i^2+n_i(\bar{y}_i-\underline{x}_i'\underline{\beta})^2]-\frac{1}{2\delta^2}(\ln\sigma_i^2-\underline{z}_i'\underline{\gamma})^2\right\},\\
&\quad \text{for} \quad i=1,2,\ldots,m.
\end{aligned}
$$

Both Corollaries can be proven easily by using Theorems 3.1 and 3.2.

3.7.3 Propriety of the posterior distributions (noninformative priors)

Theorem 3.3. Suppose we have the same assumptions as in Corollary 3.1. Then the posterior distribution of $(\underline{\beta}, \underline{\gamma}, \sigma_1^2, \ldots, \sigma_N^2, \delta^2)$ is proper.

Proof. Express the variances at single observation points x_i's in terms of $z_i'\underline{\gamma}$:

$$\sigma_i^2 = \exp(z_i'\underline{\gamma}) \quad \text{for} \quad i = m+1, m+2, \ldots, N.$$

It is easy to see that

$$\frac{1}{\prod_{i=m+1}^{N}(\sigma_i^2)^{1/2}}\exp\left[-\sum_{i=m+1}^{N}\frac{(y_i-\underline{x}_i'\underline{\beta})^2}{2\sigma_i^2}\right] = \prod_{i=m+1}^{N}\exp\left[-\frac{(y_i-\underline{x}_i'\underline{\beta})^2}{2\exp(z_i'\underline{\gamma})}-\frac{(z_i'\underline{\gamma})}{2}\right]. \tag{3.4}$$

Let $a_i = z_i'\underline{\gamma}$ for each $i = m+1, m+2, \ldots, N$. Take the logarithm of $\exp\left[-\frac{(y_i-\underline{x}_i'\underline{\beta})^2}{2\exp(z_i'\underline{\gamma})}-\frac{(z_i'\underline{\gamma})}{2}\right]$, and then take the first derivative with respect to a_i, and set it to zero:

$$(y_i-\underline{x}_i'\underline{\beta})^2\exp(-a_i)+1 \equiv 0.$$

The root of the above equation is $a_i = \ln(y_i-\underline{x}_i'\underline{\beta})^2$, at which the first derivative is 0 and the second derivative is 1. Therefore, when $a_i = \ln(y_i-\underline{x}_i'\underline{\beta})^2$ for all $i = m+1, m+2, \ldots, N$, the product in (3.4) achieves its maximum, which is

$$\exp\left[-\frac{N-m}{2} - \frac{1}{2}\sum_{i=m+1}^{N} \ln(y_i - \underline{x}_i'\underline{\beta})^2\right].$$

Denote the above maximum value by $\mathcal{C}$ and the normalizing constant before the expression in Corollary 3.1 for the joint posterior distribution as $\mathcal{N}$, then

$$\begin{aligned}
&\pi(\underline{\beta}, \underline{\gamma}, \sigma_1^2, \ldots, \sigma_N^2, \delta^2|Y) \\
\leq &\frac{\mathcal{N}\mathcal{C}}{(\delta^2)^{m/2+1}\prod_{i=1}^{m}(\sigma_i^2)^{n_i/2+1}} \exp\left[-\frac{(\underline{d} - Z\underline{\gamma})'(\underline{d} - Z\underline{\gamma}) + 2\lambda}{2\delta^2}\right] \\
&\times \exp\left[-\sum_{i=1}^{m} \frac{(n_i - 1)s_i^2 + n_i(\bar{y}_i - \underline{x}_i'\underline{\beta})^2}{2\sigma_i^2}\right].
\end{aligned} \tag{3.5}$$

The nonconstant part in the expression (3.5) is exactly of the same form as the joint posterior distribution for fully replicated design, which has been shown to be proper in Chen and Ye ([6]). □

Chapter 4

Assessment of the Bayesian Hierarchical Approach in Modelling Dual Response Surfaces

Abstract

Dual response surface methodology has become popular for off-line quality control in industrial statistical research. The major innovation of this methodology, following Taguchi's original robust parameter design idea, is to model the mean and the variance individually so they can be monitored separately. Least-squares methods and the Bayesian hierarchical method have been implemented to fit the data from different dual response design strategies. The Bayesian method is found to be more robust compared with the frequentist models under different variance scenarios. In this article, we extend our investigation of the modelling efficiencies by considering various design strategies, different numbers of control factors, and different dual response surfaces. It provides a comprehensive comparison of the modelling efficiencies between the least-squares methods and the Bayesian hierarchical method. The robustness of the Bayesian method for heavy-tailed data is also checked.

Key Words: Bayesian hierarchical model; robust parameter design; second order designs.

4.1 Introduction

The goal of the robust parameter design (RPD) methodology (Taguchi [36]) is to choose proper levels for a certain set of the control factors at which the defined process characteristic can be optimized and the sensitivity of the process to the variations in another set of uncontrollable factors is reduced. The primary motivation for the development of RPD is that the quality of a product or a process is characterized not only by the mean but also the variance.

Dual response surface methodology (RSM) is proposed by Vining and Myers ([40]) as an alternative approach to Taguchi's crossed-array designs and signal-to-noise ratios to achieve the goal of RPD. Vining and Myers ([40]) suggest that the process characteristic and its variability form a dual response system, and two separate models can be built for the response and its variance. In that paper, a second order design is replicated and two complete second order models are built for the sample mean and the sample standard deviation respectively. The model parameters are estimated with the ordinary least squares method. Vining and Myers ([40]) mention that the generalized least squares (GLS) method should be used to estimate the coefficients in the mean model in order to account for the heterogeneity of variances in dual response problems. Chen and Ye ([6]) propose a Bayesian hierarchical approach for fully replicated second order designs and have shown that the Bayesian method gives better overall performance than two frequentist least squares methods.

Vining and Schaub ([41]) suggest that often the variance function follows a first order model and the mean follows a second order model. Hence there is no need to replicate the whole second order design. Therefore, the design strategy in dual response problems includes selecting a second order design as the basic design, and choosing which portion of the design to be replicated. Vining and Schaub ([41]) use the D criterion (to maximize the determinant of the information matrix $\mathbb{M} = X'X$) to compare different second order designs and different replication strategies. The choice of the design strategy is not covered in this paper. Please refer to Vining and Schaub ([41]) for details.

The least squares methods can easily be adjusted to fit the data from partially replicated designs by considering only the replicated design points in fitting the variance model. Chen and Ye ([7]) extend application of the Bayesian method to partially replicated designs. It also investigates the modelling efficiencies of the Bayesian method and two least square methods for different replication strategies when the basic second order design is a central composite design with 3 control factors.

This paper provides a comprehensive comparison of the modelling efficiencies of the least squares methods and the Bayesian model. The performance of the Bayesian method is assessed under various design strategies, different numbers of control factors, and different dual response surfaces. With these comparisons, this paper can be used a reference for the researchers. When prior knowledge on the dual response surfaces is available, the researchers can choose the design strategy based on their requirements on the modelling efficiency and their time and budget limitations on the experiment. The paper is organized as follows. Section 4.2 reviews some popular second order designs, all of which are used as the basic designs for simulation studies in Section 4.3. After that, the simulation studies are undertaken under various situations. Section 4.4 contains summaries and conclusion.

4.2 Commonly Used Second Order Designs

In this section, some popular second order designs are reviewed. These designs plus the 3^k complete cuboidal design are used as the basic designs in the simulation studies in Section 4.3. For details about these designs, readers are referred to Myers and Montgomery ([26]) for more information.

4.2.1 Central composite designs

The central composite design (CCD), introduced by Box and Wilson ([5]), is the most popular class of second order designs. It involves the use of a *two-level factorial* or *fraction* resolution

V, $2k$ axial points, and n_c center runs. The two-level factorial or resolution V fraction contributes to the estimation of the linear terms and the two way interactions. The axial points are at a distance of α from the design center, and they are primarily used to estimate the quadratic terms. The center runs provide an estimate of the pure error and contribute toward the estimation of the quadratic terms. The areas of flexibility in the use of the central composite designs reside in the selection of α, the axial distance, and n_c, the number of center runs. The choice of α depends, to a large extent, on the region of operability and the region of interest. The value of α generally varies from 1 to $\sqrt{k}$, the former placing the axial points on the face of the cube or hypercube, the latter on a common sphere. The choice of n_c provides flexibility to get a better estimate of pure error and better power for tests.

4.2.2 Small composite designs

The small composite design (SCD), developed by Hartley ([17]), is a small economical design which exists for $k \geq 3$. It gets the name from the idea of the central composite design. The only modification to the CCD is that its factorial component is a fractional factorial design of resolution III. Therefore, in the factorial portion, linear main effect terms are aliased with two-factor interaction terms. Hartley ([17]) points out that in spite of this, all coefficients in a second order model are estimable since the linear coefficients also benefit from the axial points. However, SCD suffers considerably in efficiency for estimation of linear and interaction coefficients. It is recommended that SCD should not be an alternative to CCD unless the cost of each experimental run is high. The design matrix for $k = 3$ SCD is shown in Figure 4.1(a).

4.2.3 Hoke designs

Hoke design was proposed by Hoke ([18]) for a second order model and a cuboidal region of interest when one seeks to avoid a design point in one of the corners. They are non-orthogonal designs based on an irregular fraction of the 3^k factorial portion and a set of axial points. The Hoke designs have seven different types, $D1$ through $D7$ for any k. $D1$

x_1	x_2	x_3
-1	-1	-1
1	1	-1
1	-1	1
-1	1	1
α	0	0
$-\alpha$	0	0
0	α	0
0	$-\alpha$	0
0	0	α
0	0	$-\alpha$
0	0	0

(a)

x_1	x_2	x_3
-1	-1	-1
1	-1	-1
-1	1	-1
1	1	-1
-1	-1	1
1	-1	1
-1	1	1
-1	0	0
0	-1	0
0	0	-1
1	0	0
0	1	0
0	0	1

(b)

x_1	x_2	x_3
-1	-1	-1
1	-1	-1
-1	1	-1
1	1	-1
-1	-1	1
1	-1	1
-1	1	1
1	0	0
0	1	0
0	0	1

(c)

Figure 4.1: Design matrices for several $k = 3$ second order designs: (a) SCD, (b) Hoke $D4$, and (c) Notz design.

through $D3$ are three saturated designs found to have the smallest trace of the information matrix $\mathbf{X}'\mathbf{X}$, and thus are saturated A-optimal designs. $D4$ through $D7$ are constructed through the augmentation of additional subsets to $D1$ through $D3$. The design matrix for $k = 3$ Hoke $D4$ is shown in Figure 4.1(b). Refer to Hoke ([18]) for more details on the construction and efficiencies of $D1$ through $D7$.

4.2.4 Notz designs

The Notz design ([29]) is a minimum run D optimal design for the second order model for a cuboidal region. It may be useful when one seeks to avoid a design point in one of the corners but also wants a design with as few observations as possible to estimate all coefficients in a complete second order model. The Notz design is constructed using a subset of the 2^k factorial plus k one-factor-at-a-time axial points. Specifically, the design matrix for $k = 3$ is displayed in Figure 4.1(c).

4.3 Simulation Studies

As in the two papers by Chen and Ye ([6] and [7]), data are simulated under three different variance scenarios:

- **Constant variance**:

$$y_{ij} \sim N(\underline{x}_i'\underline{\beta}, 1), \quad \text{for} \quad i = 1, 2, \ldots, N, \quad j = 1, 2, \ldots, n_j;$$

- Nonconstant variance where the standard deviation follows a normal distribution (referred to as **Stdev**):

$$y_{ij} \sim N(\underline{x}_i'\underline{\beta}, \sigma_i^2), \quad \sigma_i \sim N(\underline{z}_i'\underline{\gamma}, \eta^2), \quad \text{for} \quad i = 1, 2, \ldots, N, \quad j = 1, 2, \ldots, n_j.$$

  In generating the standard deviation, the minimum mean standard deviation over the experimental region is set to be much larger than η, so that the simulated standard deviation has little likelihood to be negative;

- Nonconstant variance where the variance follows a Log-Normal distribution (referred to as **Logvar**):

$$y_{ij} \sim N(\underline{x}_i'\underline{\beta}, \sigma_i^2), \quad \sigma_i^2 \sim \text{Log-Normal}(\underline{z}_i'\underline{\gamma}, \eta^2), \quad \text{for} \quad i = 1, 2, \ldots, N, \quad j = 1, 2, \ldots, n_j.$$

For each scenario, $\underline{x}_i'$ contains the intercept, linear terms, quadratic terms and interactions in order, and $\underline{z}_i'$ contains the intercept, linear terms, quadratic terms and interactions, if they are present in the model. For simplicity, it is assumed that the columns of the two model matrices are known in advance. That is, all three methods have no problems of under- or over-parameterizations. The optimal operating conditions are searched for over the experimental region under the MSE criterion proposed by Lin and Tu ([24]).

4.3.1 Fixed parameters, $k = 3$, normal distributions

This paper uses the relative efficiency, proposed by Chen and Ye ([6]), as a single-number measure of performance to see how accurate a model is in choosing "good points." If the operating condition recommended by a model performs similarly to the true optimum, i.e., the relative efficiency is very near to 1, then the model is viewed as good. If the relative efficiency of a model is very small, that means the model picks a point far from the optimum. When the model parameters used to generate the data are fixed values, the distribution of relative efficiency contains complete information on the performance of the modelling method given certain dual response surfaces.

In this section, the parameters used to generate the data are the estimates from fitting the printing ink data in Chen and Ye ([6]). The experiment contains 3 control factors and the responses at a certain location are assumed to follow a normal distribution, conditional on the mean model parameters and the variance at that location. The four second order designs reviewed in Section 4.2 plus the 3^3 complete cubic design are used as the basic designs for dual response surfaces. When the variance model is of complete second order, the basic design is fully replicated; and when the variance model contains no quadratic terms, the factorial portion or the axial portion of the basic design is replicated. Moreover, as replicating the center runs is a common practice in RSM, multiple observations are always taken at the center of the design if the center point is included in the basic design. Therefore, the variance model is of linear order, the factorial or the axial portion is replicated together with the center point in the central composite design, the 3^3 complete cubic design, and the small composite design. Henceforth, a factorial replicated design refers to one of which all factorial points are replicated, regardless of the presence of the center run; and a similar definition for an axial replicated design.

Of course, selecting which part of the basic design to be replicated is not restricted by the possibilities mentioned above. There are many other choices, such as replicating half factorial points and half axial points. Every one of them is likely to be adopted as long as the choice makes the subsequent statistical analysis possible. In statistics, there are many design optimality criteria to help make the decision. In practice, a lot of factors need to be taken care of, such as the cost of the total experiment, the possibility of replicating a certain design point, and the practical importance of a design strategy. Similarly, the variance model needn't to be either purely linear or completely quadratic. Any model nested within the complete second order model is considered to be sensible in dual response problems, and usually subjective knowledge mainly determine the terms to be included in the model. However, it is impossible to enumerate and investigate all possible combinations. For simplicity, apart from the fully replicated design and the second order variance model case, we only look into the combinations where a design is either factorial or axial replicated and the variance model either contains only the linear terms or all linear and interaction terms.

For each combination of the design strategy and the variance model, the number of simulated data sets is determined such that the largest standard error of the mean efficiency does not exceed 0.005 and at least 2500 data sets should be simulated, whichever is larger.

Table 4.1 lists the mean, median and standard deviation of the relative efficiency for each modelling method when the basic design is a 3^3 complete cubic design. When the design is fully replicated and the variance model is of complete second order, all three methods perform very well in the constant variance case and in the Stdev case. If the variance is of the Logvar case, the BAYES method performs best. When the design is factorial replicated and the variance models contains only linear terms, the performances of the three methods

Table 4.1: The relative efficiencies of the three methods when the basic design is a 3^3 complete cubic design.

Rep Portion	**Variance Model**	**Method**	**Constant**		**Stdev**		**Logvar**	
			Mean (s.d.)	Median	Mean (s.d.)	Median	Mean (s.d.)	Median
full	linear + quadratic + interaction	OOLS	0.9201 (0.0959)	0.9561	0.7714 (0.1693)	0.8181	0.5423 (0.2412)	0.5500
		GOLS	0.9182 (0.0986)	0.9552	0.7695 (0.1707)	0.8156	0.5425 (0.2403)	0.5463
		BAYES	0.9084 (0.1079)	0.9484	0.7693 (0.1642)	0.8025	0.7893 (0.1834)	0.8337
factorial + center	linear	OOLS	0.8532 (0.1516)	0.9053	0.7532 (0.1847)	0.8027	0.7073 (0.2012)	0.7458
		GOLS	0.8557 (0.1533)	0.9125	0.8105 (0.1516)	0.8566	0.7980 (0.1554)	0.8359
		BAYES	0.8569 (0.1512)	0.9121	0.8112 (0.1707)	0.8688	0.8145 (0.1549)	0.8494
factorial + center	linear + interaction	OOLS	0.8471 (0.1580)	0.9035	0.6675 (0.2185)	0.6468	0.4040 (0.2201)	0.3894
		GOLS	0.8427 (0.1602)	0.8995	0.7567 (0.2071)	0.8119	0.4866 (0.1980)	0.5121
		BAYES	0.8478 (0.1573)	0.9066	0.7305 (0.1562)	0.7530	0.7811 (0.2035)	0.8513
axial + center	linear	OOLS	0.8285 (0.1752)	0.8911	0.6442 (0.2321)	0.6660	0.5434 (0.2459)	0.5378
		GOLS	0.7878 (0.2066)	0.8520	0.6622 (0.2320)	0.7085	0.6044 (0.2470)	0.6529
		BAYES	0.8019 (0.1906)	0.8598	0.7693 (0.1812)	0.8209	0.7154 (0.2080)	0.7691

Table 4.2: The relative efficiencies of the three methods when the basic design is a central composite design with the number of the control factors $k = 3$ and the axial distance $\alpha = 1.682$.

Rep Portion	Variance Model	Method	Constant		Stdev		Logvar	
			Mean (s.d.)	Median	Mean (s.d.)	Median	Mean (s.d.)	Median
full	linear + quadratic + interaction	OOLS	0.8600 (0.1469)	0.9105	0.6148 (0.1998)	0.6330	0.4652 (0.2578)	0.4443
		GOLS	0.8494 (0.1549)	0.8992	0.6267 (0.2003)	0.6539	0.5110 (0.2586)	0.2586
		BAYES	0.8499 (0.1542)	0.9016	0.5880 (0.1876)	0.5875	0.6645 (0.2268)	0.7230
factorial + center	linear	OOLS	0.7962 (0.1957)	0.8569	0.5912 (0.2328)	0.5900	0.4390 (0.2741)	0.3932
		GOLS	0.7975 (0.2001)	0.8639	0.6883 (0.2059)	0.7130	0.6025 (0.2320)	0.6173
		BAYES	0.7966 (0.1944)	0.8542	0.7111 (0.2115)	0.7436	0.7300 (0.2078)	0.7729
factorial + center	linear + interaction	OOLS	0.7908 (0.1983)	0.8511	0.5528 (0.2417)	0.5535	0.5108 (0.1925)	0.5237
		GOLS	0.7805 (0.2071)	0.8400	0.6906 (0.2064)	0.7249	0.5564 (0.1789)	0.5609
		BAYES	0.7897 (0.2012)	0.8550	0.6693 (0.2157)	0.6921	0.6800 (0.1979)	0.6888
axial + center	linear	OOLS	0.7732 (0.2083)	0.8300	0.5248 (0.2580)	0.5070	0.4927 (0.2829)	0.4712
		GOLS	0.7710 (0.2114)	0.8329	0.5378 (0.2621)	0.5290	0.5474 (0.2739)	0.5573
		BAYES	0.7751 (0.2077)	0.8393	0.5264 (0.2398)	0.5057	0.6123 (0.2510)	0.6279

Table 4.3: The relative efficiencies of the three methods when the basic design is a small composite design with the number of the control factors $k = 3$ and the axial distance $\alpha = 1.682$.

Rep Portion	**Variance Model**	**Method**	**Constant**		**Stdev**		**Logvar**	
			Mean (s.d.)	Median	Mean (s.d.)	Median	Mean (s.d.)	Median
full	linear + quadratic + interaction	OOLS	0.8022 (0.1910)	0.8633	0.5172 (0.2214)	0.5380	0.4751 (0.2932)	0.4689
		GOLS	0.7977 (0.1970)	0.8651	0.5308 (0.2206)	0.5574	0.4908 (0.2965)	0.4937
		BAYES	0.7929 (0.1977)	0.8521	0.5291 (0.2154)	0.5470	0.5337 (0.2708)	0.5610
factorial + center	linear	OOLS	0.7006 (0.2504)	0.7509	0.3920 (0.2438)	0.3392	0.3545 (0.2470)	0.2851
		GOLS	0.6949 (0.1254)	0.7476	0.3972 (0.2419)	0.3536	0.3669 (0.2544)	0.2969
		BAYES	0.7069 (0.2449)	0.7617	0.4228 (0.2492)	0.3909	0.4687 (0.2781)	0.4567
axial + center	linear	OOLS	0.7308 (0.2375)	0.7938	0.4060 (0.2463)	0.3582	0.4818 (0.2629)	0.4533
		GOLS	0.7241 (0.2374)	0.7844	0.4188 (0.2483)	0.3701	0.4831 (0.2627)	0.4583
		BAYES	0.7220 (0.2393)	0.7807	0.4001 (0.2351)	0.3609	0.5449 (0.2531)	0.5520

Table 4.4: The relative efficiencies of the three methods when the basic design is a Hoke design with the number of the control factors $k = 3$.

Rep Portion	**Variance Model**	**Method**	**Constant**		**Stdev**		**Logvar**	
			Mean (s.d.)	Median	Mean (s.d.)	Median	Mean (s.d.)	Median
full	linear + quadratic + interaction	OOLS	0.8399 (0.1618)	0.8932	0.6999 (0.2137)	0.7477	0.4786 (0.2046)	0.4684
		GOLS	0.8270 (0.1726)	0.8865	0.6987 (0.2412)	0.7496	0.4873 (0.2087)	0.4878
		BAYES	0.8284 (0.1710)	0.8815	0.6775 (0.2108)	0.6985	0.6495 (0.2490)	0.7027
factorial	linear	OOLS	0.7534 (0.2248)	0.8164	0.7070 (0.2042)	0.7497	0.6336 (0.2202)	0.6677
		GOLS	0.7348 (0.2346)	0.7880	0.7432 (0.1887)	0.7954	0.6730 (0.2047)	0.7113
		BAYES	0.7609 (0.2175)	0.8180	0.7395 (0.1952)	0.7904	0.7431 (0.1845)	0.7838
axial	linear	OOLS	0.7464 (0.2294)	0.8136	0.6269 (0.2372)	0.6511	0.5484 (0.2496)	0.5604
		GOLS	0.7462 (0.2304)	0.8076	0.6247 (0.2354)	0.6406	0.5458 (0.2497)	0.5550
		BAYES	0.7422 (0.2281)	0.7967	0.6982 (0.2133)	0.7416	0.5959 (0.2397)	0.6147

Table 4.5: The relative efficiencies of the three methods when the basic design is a Notz design with the number of the control factors $k = 3$.

Rep Portion	**Variance Model**	**Method**	**Constant**		**Stdev**		**Logvar**	
			Mean (s.d.)	Median	Mean (s.d.)	Median	Mean (s.d.)	Median
full	linear + interaction	OOLS	0.8118 (0.1881)	0.8738	0.6911 (0.2341)	0.6712	0.6218 (0.1515)	0.6426
		GOLS	0.8118 (0.1881)	0.8738	0.6911 (0.2341)	0.6712	0.6218 (0.1515)	0.6426
		BAYES	0.8100 (0.1885)	0.8680	0.6686 (0.1921)	0.6889	0.6760 (0.1800)	0.6664
factorial	linear	OOLS	0.7348 (0.2346)	0.7880	0.6434 (0.2431)	0.6858	0.5816 (0.2356)	0.6088
		GOLS	0.7348 (0.2346)	0.7880	0.6434 (0.2431)	0.6858	0.5816 (0.2356)	0.6088
		BAYES	0.7354 (0.2378)	0.7965	0.6574 (0.2420)	0.7076	0.6062 (0.2419)	0.6312

Table 4.6: The relative efficiencies of the three methods when the basic design is a central composite design with the number of the control factors $k = 4$ and the axial distance $\alpha = 2$.

Rep Portion	**Variance Model**	**Method**	**Constant**		**Stdev**		**Logvar**	
			Mean (s.d.)	Median	Mean (s.d.)	Median	Mean (s.d.)	Median
full	linear + quadratic + interaction	OOLS	0.8666 (0.1442)	0.9205	0.3953 (0.2201)	0.3516	0.2146 (0.2041)	0.1387
		GOLS	0.8527 (0.1537)	0.9067	0.4230 (0.2278)	0.3759	0.3264 (0.2256)	0.2887
		BAYES	0.8565 (0.1524)	0.9119	0.3703 (0.2028)	0.3094	0.5513 (0.2267)	0.5238
factorial + center	linear	OOLS	0.8223 (0.1765)	0.8827	0.5983 (0.2073)	0.6032	0.5317 (0.2400)	0.5241
		GOLS	0.8197 (0.1830)	0.8838	0.6840 (0.1851)	0.7007	0.7099 (0.1860)	0.7497
		BAYES	0.8183 (0.1788)	0.8751	0.7222 (0.1824)	0.7551	0.7700 (0.1496)	0.7948
factorial + center	linear + interaction	OOLS	0.8272 (0.1725)	0.8862	0.2594 (0.2126)	0.1904	0.2886 (0.2237)	0.1998
		GOLS	0.8131 (0.1834)	0.8710	0.3265 (0.2264)	0.2742	0.4144 (0.2512)	0.3704
		BAYES	0.8210 (0.1797)	0.8842	0.3382 (0.2038)	0.3081	0.5836 (0.2783)	0.6438
axial + center	linear	OOLS	0.7824 (0.2090)	0.8486	0.4392 (0.2376)	0.4035	0.4077 (0.2357)	0.3894
		GOLS	0.7629 (0.2178)	0.8247	0.4604 (0.2397)	0.4319	0.4740 (0.2350)	0.4682
		BAYES	0.7660 (0.2167)	0.8287	0.4534 (0.1742)	0.4233	0.6415 (0.2052)	0.6608

Table 4.7: The relative efficiencies of the three methods when the response comes from a t_4 distribution, and the basic design is a central composite design with the number of the control factors $k = 3$ and the axial distance $\alpha = 1.682$.

Rep Portion	**Variance Model**	**Method**	**Constant**		**Stdev**		**Logvar**	
			Mean (s.d.)	Median	Mean (s.d.)	Median	Mean (s.d.)	Median
full	linear + quadratic + interaction	OOLS	0.8237 (0.1764)	0.8822	0.5700 (0.2157)	0.5908	0.3705 (0.2624)	0.3117
		GOLS	0.8212 (0.1796)	0.8763	0.5880 (0.2097)	0.6067	0.4340 (0.2684)	0.4088
		BAYES	0.8291 (0.1783)	0.8957	0.5679 (0.1958)	0.5663	0.5959 (0.2610)	0.6520
factorial + center	linear	OOLS	0.7319 (0.2446)	0.7994	0.5099 (0.2589)	0.4870	0.3536 (0.2732)	0.2822
		GOLS	0.7291 (0.2485)	0.7997	0.5966 (0.2424)	0.6052	0.4948 (0.2574)	0.4868
		BAYES	0.7393 (0.2414)	0.8105	0.6437 (0.2455)	0.6749	0.6490 (0.2439)	0.6824
factorial + center	linear + interaction	OOLS	0.7423 (0.2376)	0.8101	0.4959 (0.2625)	0.4802	0.4799 (0.2032)	0.4917
		GOLS	0.7425 (0.2368)	0.8105	0.6146 (0.2354)	0.6357	0.5292 (0.1904)	0.5407
		BAYES	0.7568 (0.2223)	0.8318	0.6181 (0.2415)	0.6493	0.6427 (0.2093)	0.6492
axial + center	linear	OOLS	0.7151 (0.2493)	0.7761	0.4453 (0.2634)	0.4024	0.4881 (0.2808)	0.4775
		GOLS	0.7158 (0.2497)	0.7777	0.4636 (0.2691)	0.4248	0.5469 (0.2725)	0.5673
		BAYES	0.7359 (0.2393)	0.8060	0.4635 (0.2560)	0.4311	0.6216 (0.2495)	0.6413

are similar in the constant variance case. In two nonconstant variance cases, the BAYES method is much better than OOLS and slightly better than GOLS. When the design is factorial replicated and the variance model contains the linear and interaction terms, the results show that the BAYES model is a little worse than GOLS when the variance is Stdev; and it completely dominates in terms of the mean and median of the relative efficiencies in the Logvar case. The performances of the three methods are a bit strange when the design is axial replicated and the variance model contains only the linear terms: the relative efficiencies of OOLS and GOLS are significantly lower than the BAYES method in the Stdev case, although they have no model misspecification while the BAYES does have. A possible explanation for this phenomenon is that replicating only the axial points is not a good decision for OOLS and GOLS while the BAYES method is quite robust to the choice of the design strategy.

Readers may have noticed that in some combinations of the design strategy and the variance model, the relative efficiencies of BAYES in the Stdev case are even higher than those in the Logvar case. It may appear uncommon since the BAYES method misspecifies the variance model in the Stdev case while not in the Logvar case. However, it is unfair to compare the performance across different variance scenarios, as the true dual response surfaces are different and it also results in different degrees of the difficulty to approximate the true response.

We can also make comparisons across the two combinations: the factorial replicated design with the linear variance model, and the axial replicated design with the linear variance model, since the true dual responses are identical. It is observed that all three methods perform much worse in the axial replicated design than in the factorial replicated design. One reason is that the axial replicated design does not have as many support points for the variance model as in the factorial design, and the other reason is that the axial points are not as disperse as the factorial points to cover the region of interest. In Vining and Schaub ([41]), the axial replicated design strategy is not recommended based on the D-optimal criterion.

Table 4.2 lists the mean, median and standard deviation of the relative efficiencies when the basic design is a central composite design with $k = 3$ control factors and the axial distance

α equal to 1.682. As discussed for the CCD in Section 4.2, the axial distance can be any real number between 1 and $\sqrt{k}$. In this paper only the CCD with the axial distance equal to 1.682 is looked into. Except for the two classes of composite designs, CCD and SCD, all other designs are cubic. We choose to make the two composite designs to be spherical so that we have a variety of the regions of interest. For the listed four combinations, the performance of the Bayes method is fairly comparable to the two least squares in the constant variance case and the Stdev case. In the Logvar case, the performances of OOLS and GOLS are much worse due to model misspecification. By comparing Table 4.1 (3^3 cubic design) and Table 4.2 (CCD), we observe that for the same combination of the design strategy and the variance model, the relative efficiency in the CCD is always smaller than that in the 3^3 cubic design. This is because the CCD uses fewer experimental runs than the 3^3 cubic design, while the region of interest in the CCD is a sphere, which has more volume than the cubic in the 3^3 cubic design.

Table 4.3 is for the combination where the basic design is a small composite design with 3 control factors and the axial distance α equal to 1.682. Not like Tables 4.1 and 4.2, Table 4.3 does not have the combination where the design is factorial replicated and the variance model contains the linear terms and the interaction terms. The reason is replicating a resolution *III* fraction of the factorial portion and the center run does not support the estimation of a variance model containing both the linear terms and interaction terms. The comparison of the relative efficiency for the SCD is much alike that for the CCD. Compared to the CCD, the SCD has lower relative efficiencies. This is expected since the SCD has fewer experimental runs than the CCD. Based on the modelling efficiency, the SCD is also not recommended, which is consistent with those using design optimality criteria.

Table 4.4 lists the mean, median and standard deviation of the relative efficiency when the basic design is a Hoke design with 3 control factors. Similar to the SCD, the Hoke design does not have the combination of replicating the factorial portion and the variance model with the linear terms and the interactions. The reason for lack of this combination is slightly different: the 7 replicated factorial points can support the estimation of the variance model using either OOLS and GOLS, but not enough for using the BAYES method. As discussed in Chen and Ye ([7]), the variance model in the BAYES method has one more parameter compared with OOLS and GOLS. Therefore, using the BAYES method to estimate the variance model in this case requires at least 8 replicated points. The relative efficiency comparison for Table 4.4 is similar to Table 4.2. The performance of the BAYES method is robust to model misspecification in the Stdev case and still dominates in the Logvar case. Particularly, OOLS and GOLS are worse than the BAYES model in the Stdev case when the design is axial replicated and the variance model contains only the linear terms. This can be explained in the same way as for the 3^3 cubic design.

Table 4.5 contains information on the modelling efficiency when the basic design is a Notz design with 3 control factors. The Notz design is saturated for a second order model if the least squares methods are used for estimation. However, the BAYES method needs at least 11 replicated points to estimate a complete second order variance model. Therefore, when the Notz design is fully replicated, the variance model is assumed to contain no quadratic terms. When the Notz design is factorial replicated, the variance model contains only the linear terms because of the same reason as stated in the Hoke design. From Table 4.5, we cannot distinguish any method from the other two. The speciality about the Notz design is the equivalence between OOLS and GOLS: the estimates of the model parameters, the optimization results, and the relative efficiencies are identical regardless of the design strategy and the variance model. This speciality is not only limited to the Notz design. It is shown below that as long as the basic design is saturated for the mean response model, OOLS and GOLS are always equivalent.

Let X be the model matrix for the mean response, and it is square and of full rank if the basic design is saturated for the mean model. The variance model parameters $\underline{\gamma}$ are estimated in the same way in OOLS and GOLS, so we only need to examine if the mean model parameters $\underline{\beta}$ are equivalent. OOLS estimates $\underline{\beta}$ as

$$\underline{\hat{\beta}}_{OOLS} = (X'X)^{-1}X'\underline{\bar{y}} = X^{-1}(X')^{-1}X'\underline{\bar{y}} = X^{-1}\underline{\bar{y}},$$

and GOLS estimates them as

$$\underline{\hat{\beta}}_{GOLS} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\underline{\bar{y}} = X^{-1}\hat{V}(X')^{-1}X'\hat{V}^{-1}\underline{\bar{y}} = X^{-1}\underline{\bar{y}},$$

where $\hat{V}^{-1}$ is the predicted variance matrix based on the variance model, and $\underline{\bar{y}}$ is the vector of the sample means. Therefore both $\underline{\hat{\beta}}$ and $\underline{\hat{\gamma}}$ are identical, and so are the optimization results and modelling efficiencies, since the subsequent analysis only depends on $\underline{\hat{\beta}}$ and $\underline{\hat{\gamma}}$.

4.3.2 Fixed parameters, $k = 4$, normal distributions

The simulation setting in this section is very similar to Section 4.3.1 except that the number of control factors in the design is 4, instead of 3. We would not investigate each class of the basic design as in Section 4.3.1. The CCD is used as the basic design in this section as it is the most popular class of designs for second order models.

Table 4.6 lists some summary statistics of the relative efficiencies when the number of the control factors $k = 4$ and the axial distance of the CCD is $\alpha = 2$. From the table we can see that the dual response surfaces with 4 control factors are usually more difficult to model than those with 3 control factors, especially when the variance surface becomes complicated. The advantage of GOLS over OOLS is much more obvious for 4 control factors designs. Within each combination of the design strategy and the variance model, the BAYES method still displays more robustness: in the Stdev case BAYES is similar to the least squares methods though with model misspecification; while OOLS and GOLS suffer from model misspecification in the LOGVAR case, and their performance drops dramatically.

4.3.3 Fixed parameters, $k = 3$, t_4 distribution

This section uses simulated data to investigate the modelling efficiency when the data come from a heavy tailed distribution. The basic design is a CCD with 3 control factors. In simulation, the response at the design point $\underline{x}_i$ is assumed to follow a t_4 distribution conditional on the mean model parameters $\underline{\beta}$ and the variance σ_i^2.

The relative efficiencies of the three methods are listed in Table 4.7. Compared to Table 4.2, Table 4.7 has lower modelling efficiencies due to the non-normality of the data. The disadvantage of the two least square methods is very obvious here. Among the four combinations of the design strategy and the variance model, the median and mean of OOLS' and GOLS's relative efficiencies exceeds 0.5 in the Logvar case only when the design is factorial replicated and the variance model contains merely the linear terms, where the information provided by the replicated points is fairly abundant to estimate a linear variance model. On the other hand, the performance of BAYES is still good in the Stdev case.

If we compare across the variance scenarios within each modelling method, we find that in some combinations of the design strategy and the variance model, the relative efficiencies of OOLS and GOLS in the Stdev case are even lower than those in the Logvar case. It may appear uncommon since the least squares methods have model misspecification in the Logvar case while not in the Stdev case. As explained for the similar phenomenon in Table 4.1, it is inappropriate to compare the performance across different variance scenarios. Different true dual response surfaces lead to different degrees of difficulty in modelling the true response surfaces.

4.3.4 Random parameters, $k = 3$, normal distributions

As in Sections 4.3.2 and 4.3.3, the CCD is also used as the basic design in this section. The major difference between this section and the previous three sections is that the parameters used to simulate the data are no longer fixed here. Fixed model parameters indicate fixed dual response surfaces, and thus the fixed level of difficulty for modelling. Therefore, the simulation studies in the previous three sections are restricted to certain fixed dual response surfaces, and the summary statistics of the relative efficiency tells us the overall performance of each method conditional on the specific dual response surfaces. In this section, the model parameters are generated from a multivariate normal distribution before a data set is simulated. The fixed parameters used in Section 4.3.1 are used as the mean vector of the multivariate normal distribution. The covariance matrix is assumed to be diagonal, and the diagonal elements are set equal to the absolute values of their corresponding means. In this way, the simulated parameter values are not squeezed around their means and the true dual response surfaces are greatly different. By comparing the three methods under different surfaces, we can assess their performances in a larger scale.

Since each data set is generated from different parameters, we need to compare the relative efficiencies for the same data. Summarizing the relative efficiencies over different model parameters with the mean or median ignores the different difficulty in modelling different surfaces. Therefore we use the ratio of the relative efficiencies as a measure to compare how good the three methods are in modelling the same data set. The relative efficiency of the BAYES model is used as the baseline and it is compared against the other two methods, with "OOLS/BAYES" being the ratio of the relative efficiency of OOLS against BAYES, and "GOLS/BAYES" the ratio of GOLS against BAYES. If the ratio is larger than 1, it indicates that OOLS or GOLS is better than BAYES for that specific data set, and vice versa. Figures 4.2 - 4.5 display the boxplots for the logarithm of the ratios of the relative efficiencies. The logarithm of the ratios are plotted so that the ratios below "1" can stretch down in the same scale as those above "1" stretch upside. The box itself contains the middle half of the logarithm of the efficiency ratio. The upper hinge of the boxplot indicates the 75^{th} percentile, and the lower hinge indicates the 25^{th} percentile. The ends of the whiskers indicate the minimum and the maximum, unless outliers are present in which case the whiskers extend to a maximum of the 1.5 times the inter-quartile range.

From figures 4.2 - 4.5, it is observed that in the constant variance case, the three methods perform similarly since the major parts of the first pair of boxplots in each figure are centered at the zero line and the upper parts and lowers parts are symmetric around the line. When the variance is of the Stdev case, BAYES is as good as GOLS and slightly better than OOLS when the design is partially replicated and the variance follows a first order model (see Figures 4.3 - 4.5); and BAYES is significantly better than the other two in the fully replicated design and second order variance model scenario (see Figure 4.2). When the variance follows a Log-Normal distribution, BAYES always dominates.

Sections 4.3.1 - 4.3.4 compare the modelling efficiencies by varying the design strategy, the variance function, the number of control factors, the distribution of the measurements, and the true response surfaces. These comparisons lead to a common conclusion: the BAYES method is very robust to model misspecification while the two least squares methods are not. In certain circumstance, the BAYES method is even more efficient than the least squares

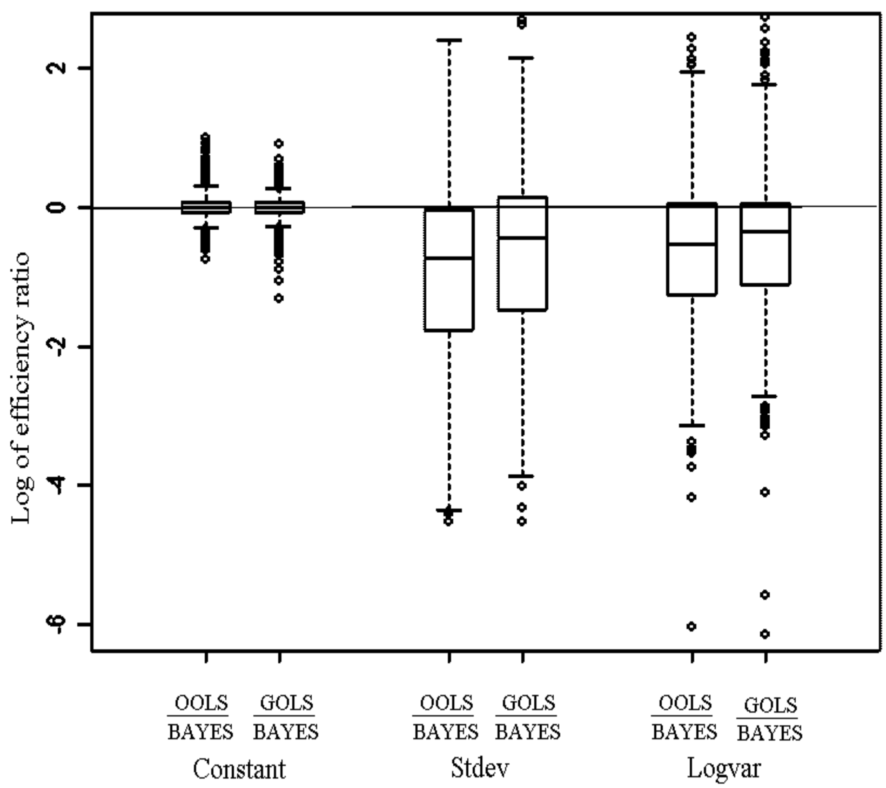

Figure 4.2: Boxplots for the logarithm of the relative efficiency ratios when the CCD is fully replicated and the variance model is of complete second order. Model parameters are simulated from multivariate normal distributions.

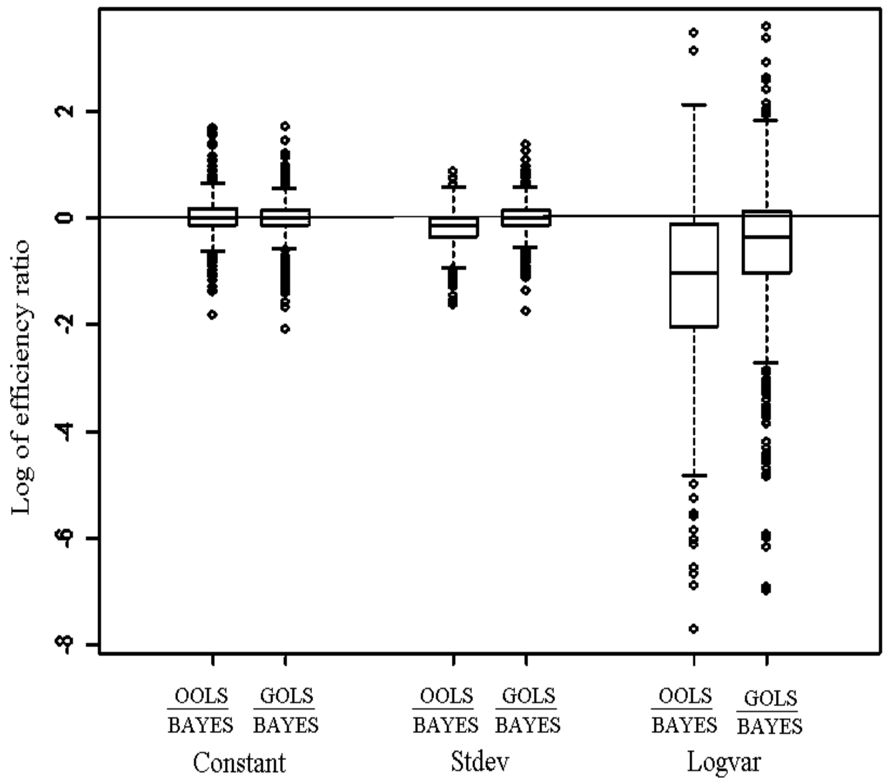

Figure 4.3: Boxplots for the logarithm of the relative efficiency ratios when the CCD is factorial replicated and the variance model contains linear terms. Model parameters are simulated from multivariate normal distributions.

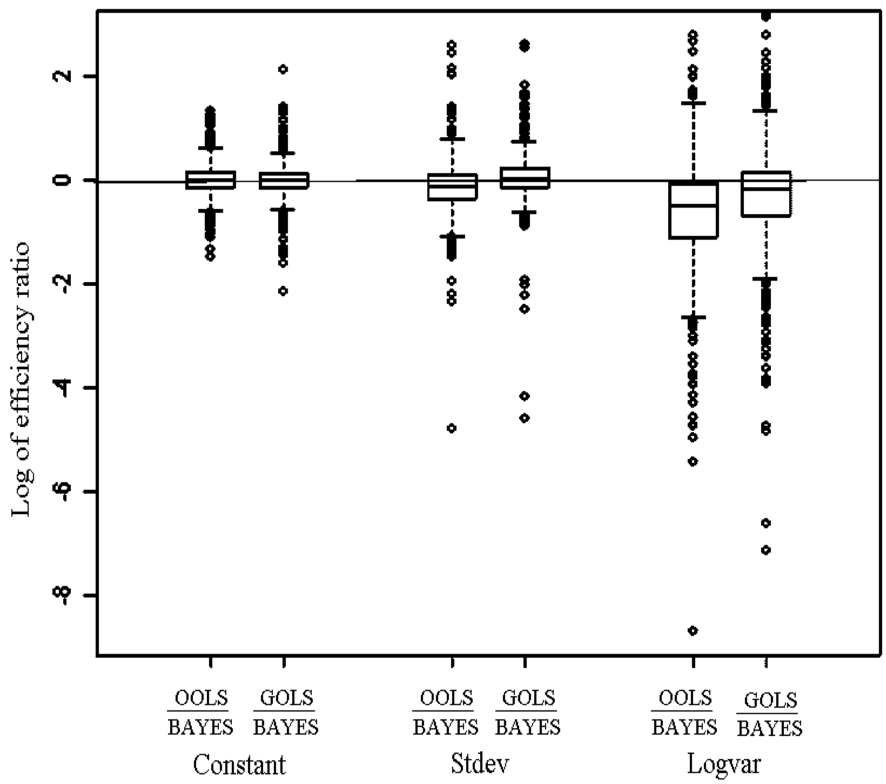

Figure 4.4: Boxplots for the logarithm of the relative efficiency ratios when the CCD is factorial replicated and the variance model contains linear and interaction terms. Model parameters are simulated from multivariate normal distributions.

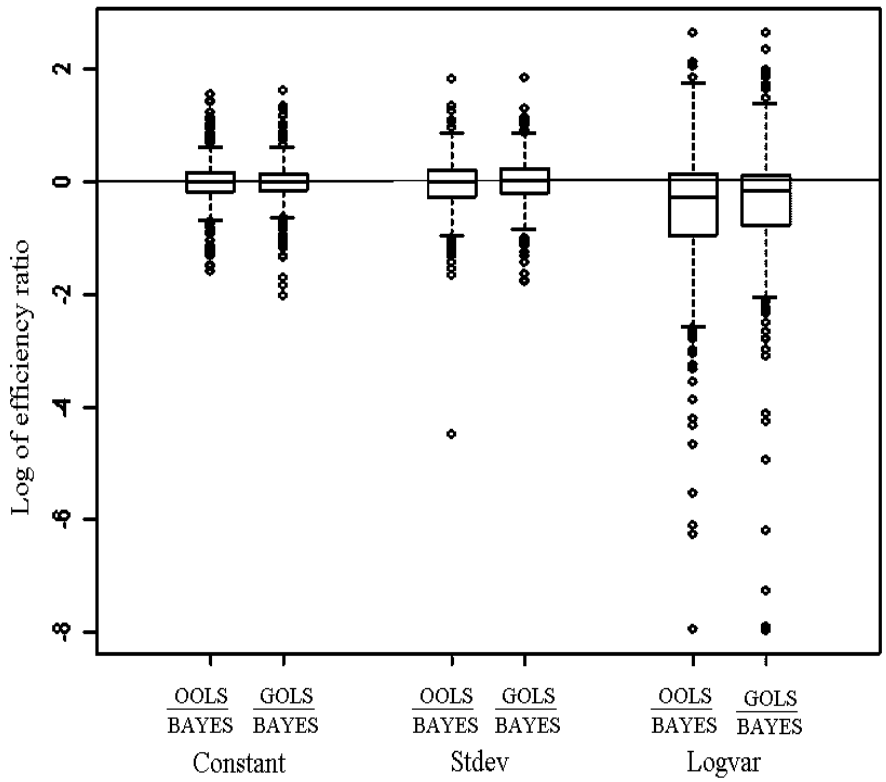

Figure 4.5: Boxplots for the logarithm of the relative efficiency ratios when the CCD is axial replicated and the variance model contains linear terms. Model parameters are simulated from multivariate normal distributions.

methods in estimating models based on poor design strategies, despite model misspecification.

4.4 Conclusions

This paper offers a comprehensive investigation on the modelling efficiencies of the Bayesian model and two least squares methods under very different situations. The Bayesian method is found to be robust to the true variance surface as well as the design strategy. The robustness to the variance surface allows the researchers to establish models with less concern about the nature of the data. They can simply choose the Bayesian method and still be pretty sure that performance at the predicted optimal operating condition based on the Bayesian model will be overall quite satisfying. And the robustness to the design strategy offers flexibility in choosing the replicated design points if a certain point is hard or impossible to be replicated due to practical constraints.

Apart from the sweeping comparison, this paper can also be used a reference for the researchers. When prior knowledge on the dual response surfaces is available, the researchers can choose the design strategy based on their requirements on the modelling efficiency and their time and budget limitations on the experiment. Bear in mind that the Bayesian model requires one more replicated point than the least squares methods. If only a saturated design can be afforded, the researcher has to go back to the least squares methods.

Chapter 5

Bayesian Methodology for Multivariate Dual Response Surfaces

5.1 Introduction

Multiple response surface methods have proven useful for improving the quality of products and processes. For most products, quality is multidimensional, so it is common to observe multiple responses on the experimental units. Many applications of multiple response surface methods can be found in the literature, including books by Khuri and Cornell ([21]) and Myers and Montgomery ([26]). When dealing with multiple response data, analyzing each response separately is not satisfactory, especially when the multiple responses are correlated. One simple way of dealing with multiple response surfaces when there are only a small number of factors is to create overlapping contour plots of the response surfaces. They may lead to an area in the experimental region where each of the mean responses satisfies the requirements or where a compromise can be obtained. Derringer and Suich ([13]) proposed to use desirability functions which turn the multiple response problem into a single response problem. Other papers account for a common covariance matrix for the multiple responses.

Chiao and Hamada ([8]) consider the covariance matrix to be dependent on the control factors as suggested by Pignatiello ([31]). It can be viewed as an extension of the dual response surface methodology. In the paper, they model the parameters of the multiple response distribution in terms of the experimental factors as follows. Let x denote a matrix of covariates which includes an intercept, main effects, quadratic terms, and two-way interactions associated with the control factors in the experiment. The distributional parameters of the m dimensional multivariate responses are modelled as:

$$\begin{aligned} \mu_k &= x\alpha_k, && k = 1, \ldots, m, \\ \log(\sigma_k^2) &= x\beta_k, && k = 1, \ldots, m, \\ \tanh^{-1}(\rho_{kl}) &= x\gamma_{kl}, && 1 \leq k < l \leq m, \end{aligned} \tag{5.1}$$

where α_k, β_k, and γ_k are column vectors, and their nonzero components determine what variables in the matrix x are included in the respective models. The inverse hyperbolic tangent transformation (Rao [32]), which is defined as:

$$\tanh^{-1}(\rho) = \frac{1}{2} \log \left(\frac{1+\rho}{1-\rho} \right),$$

is used to model the correlation. The explanation for using $\tanh^{-1}(\rho)$ is that $0 \leq (1+\rho)/2 \leq 1$ is a proportion, and taking the logit transformation of a proportion is a common practice. In Chiao and Hamada's paper ([8]), the sample means, variances, and correlations are used as the responses for the above models, and ordinary least squares methods are applied to estimate the distributional parameters. After that the levels of control factors are determined to optimize the overall performance of the multiple responses under a suitable criterion.

Apart from the overlapping contour plots and the desirability functions mentioned above, there are many other optimization criteria proposed in the literature to address different needs in multiple response surface systems. Del Castillo ([11]) suggests a constrained confidence region approach to multiresponse optimization. Chiao and Hamada ([8]) adopt the proportion of conformance criterion to maximize the probability that all responses simultaneously meet their respective specifications. Khuri and Conlon ([22]), Pignatiello ([31]), Ames *et al.* ([2]), and Vining ([39]) propose various types of quadratic loss functions to quantify the distance between the predicted multiresponse performance and the required target values.

5.2 Extension of the Bayesian Hierarchical Model

Chiao and Hamada's approach allows the covariance matrices of the multiple responses to be dependent on the setting of the control factors. By modelling the means, variances, and correlations with different models, we can monitor the three characteristics of a multiple response problem at the same time. One disadvantage on Chiao and Hamada's approach is in the estimation of the parameters for the mean model. Using the ordinary least squares method ignores the non-constant correlations as well as the non-constant variances in multiple dual response problems.

The Bayesian hierarchical approach can easily accommodate this by incorporating the nonconstant variances and correlations into the probability distribution of the data. As in the Bayesian model for univariate dual response surfaces, we first need to set the conditional distributions of the measurements and the covariance matrices. The equations in 5.1 can be easily adjusted to get the conditional distributions:

$$
\begin{aligned}
&y_k \sim MVN(x\alpha_k, \sigma_k^2), && k = 1, \ldots, m, \\
&\log(\sigma_k^2) \sim N(x\beta_k, \eta_k^2), && k = 1, \ldots, m, \\
&\tanh^{-1}(\rho_{kl}) \sim N(x\gamma_{kl}, \eta_\rho^2), && 1 \leq k < l \leq m.
\end{aligned} \tag{5.2}
$$

The distributions for the hyperparameters α_k, β_k, γ_k, η_k^2, and η_ρ^2 are specified depending on the availability of historical information.

After model building, we can derive the joint and conditional posterior distributions, use the Gibbs sampling to obtain the marginal posterior distributions for each parameter, and then compute the posterior medians or means for the point estimates for the model parameters.

5.3 An Example

The data from a plasma enhanced chemical vapor deposition experiment is used to illustrate the application of the Bayesian hierarchical model for multivariate dual response surfaces.

The experiment, first reported by Tong and Su ([38]), is designed to improve a plasma enhanced vapor deposition (PECVD) process in the fabrication of integrated circuits. In the original experiment, eight control factors are included. An $\mathcal{L}_{18}$ orthogonal array is used to accommodate one two-level factor and seven three-level factors, and each of the 18 runs is replicated five times (Taguchi [36]). The two quality characteristics of interest are the deposition thickness (Y_1) in Angstroms and a refractive index (Y_2).

The common practice in DRS is to assume a complete second order model for the mean response. An 18-run experiment cannot support a second order mean response model. For illustration purpose, we choose three three-level control factors out of the eight, so that a second order mean response model can be estimated. The three selected control factors are: chamber temperature (x_1), flow rate of SiH_4 (x_2), and chamber pressure (x_3). These three factors are selected at random, not based on the statistical significance or practical importance of the factors. The levels of the three selected factors and bivariate response data are displayed in Table 5.1.

Table 5.1: Partial data from the chemical vapor deposition experiment

Run	x_1	x_2	x_3	Y	Replicates					$\bar{y}$	s	ρ
1	-1	-1	-1	Y_1	694	839	728	688	704	730.6	62.48	-0.906
				Y_2	2.118	1.919	1.985	2.085	2.056	2.0326	0.0802	
2	-1	0	0	Y_1	918	867	861	874	851	874.2	25.89	-0.527
				Y_2	2.205	2.240	2.234	2.165	2.275	2.2238	0.0412	
3	-1	1	1	Y_1	936	954	930	1058	958	967.2	52.11	-0.933
				Y_2	2.677	2.643	2.714	2.456	2.565	2.6110	0.1026	
4	0	-1	0	Y_1	765	828	842	768	801	800.8	34.62	-0.925
				Y_2	2.096	1.997	1.949	2.046	2.000	2.0176	0.0557	
5	0	0	1	Y_1	709	743	753	752	989	789.2	113.12	-0.937
				Y_2	2.032	2.007	1.943	2.003	1.845	1.9660	0.0751	
6	0	1	-1	Y_1	795	785	846	722	833	796.2	48.65	-0.840
				Y_2	1.860	1.838	1.842	1.999	1.858	1.8794	0.0675	

Table 5.1: Partial data from the chemical vapor deposition experiment (cont')

Run	x_1	x_2	x_3	Y	Replicates					$\bar{y}$	s	ρ
7	1	0	1	Y_1	711	816	1085	787	1150	909.6	194.80	-0.945
				Y_2	2.012	1.909	1.797	1.930	1.819	1.8934	0.0873	
8	1	1	-1	Y_1	580	644	602	607	811	648.8	93.55	-0.654
				Y_2	1.834	1.760	1.760	1.782	1.744	1.7760	0.0351	
9	1	-1	0	Y_1	590	812	627	595	609	646.6	93.57	0.1743
				Y_2	1.719	1.707	1.676	1.704	1.675	1.6962	0.0197	
10	-1	1	0	Y_1	917	1142	1126	916	966	1013.4	112.08	-0.8716
				Y_2	2.097	1.911	1.889	2.014	1.960	1.9742	0.0838	
11	-1	-1	1	Y_1	1389	1405	1219	2063	1392	1493.6	327.39	-0.987
				Y_2	1.927	1.860	1.945	1.539	1.867	1.8276	0.1655	
12	-1	0	-1	Y_1	865	914	993	838	893	900.6	59.08	-0.913
				Y_2	1.963	1.881	1.812	1.923	1.899	1.8956	0.0559	
13	0	0	-1	Y_1	827	884	884	851	1066	902.4	94.56	-0.802
				Y_2	1.903	1.829	1.788	1.863	1.767	1.8300	0.0551	
14	0	1	0	Y_1	787	805	780	776	976	824.8	85.25	-0.714
				Y_2	2.103	2.020	2.011	2.107	1.968	2.0418	0.0610	
15	0	-1	1	Y_1	739	779	745	724	976	792.6	104.48	-0.872
				Y_2	2.182	2.080	2.071	2.179	1.968	2.0960	0.0888	
16	1	1	1	Y_1	724	721	690	1023	915	814.6	146.63	-0.683
				Y_2	2.274	2.166	2.215	2.103	2.203	2.1922	0.0632	
17	1	-1	-1	Y_1	771	806	785	869	859	818.0	43.94	-0.522
				Y_2	1.942	1.905	1.909	1.916	1.900	2.9144	0.0165	
18	1	0	0	Y_1	712	781	749	692	760	738.8	36.20	-0.981
				Y_2	2.077	1.961	1.985	2.101	1.980	2.0208	0.0635	

Due to the absence of subjective knowledge on the model parameters, noninformative priors are assigned to hyperparameters. Uniform priors are used for coefficients in the mean, variance and correlation models. For the variance terms in the variance and correlation models, the following priors are used:

$$\pi(\eta_1^2) \propto \frac{1}{\eta_1^2}\exp(-\frac{\lambda_1}{\eta_1^2}), \quad \pi(\eta_2^2) \propto \frac{1}{\eta_2^2}\exp(-\frac{\lambda_2}{\eta_2^2}), \quad \pi(\eta_\rho^2) \propto \frac{1}{\eta_\rho^2}\exp(-\frac{\lambda_\rho}{\eta_\rho^2}),$$

where δ_1^2, δ_2^2, and δ_ρ^2 are the variance terms as defined in (5.2), and λ_1, λ_2, and λ_ρ are small positive constants. Estimating the model coefficients with posterior medians, the fitted models are as follows:

$$\begin{aligned}
\hat{y}_1 &= 833 - 70.4x_1 + 5.7x_2 + 18.1x_3 - 17.0x_1^2 - 37.0x_2^2 + 35.0x_3^2 \\
&\quad -65.1x_1x_2 - 16.8x_1x_3 - 1.1x_2x_3, \\
\hat{y}_2 &= 2.08 - 0.11x_1 + 0.07x_2 + 0.12x_3 - 0.05x_1^2 - 0.07x_2^2 - 0.02x_3^2 \\
&\quad +0.07x_1x_2 - 0.12x_1x_3 + 0.21x_2x_3, \\
\ln\hat{\sigma}_1^2 &= 7.54 + 0.20x_1 - 0.03x_2 + 0.72x_3 + 0.61x_1^2 + 0.65x_2^2 + 0.94x_3^2 \\
&\quad -0.11x_1x_2 + 0.43_1x_3 - 0.61x_2x_3, \\
\ln\hat{\sigma}_2^2 &= -4.49 - 0.71x_1 - 0.20x_2 + 0.74_3 - 0.77x_1^2 - 0.36x_2^2 + 0.26x_3^2 \\
&\quad +0.39x_1x_2 + 0.03_1x_3 - 0.16x_2x_3, \\
\tanh^{-1}(\rho_{12}) &= -0.31 - 0.05x_1 - 0.10x_2 + 0.13x_3.
\end{aligned}$$

The target values are 1000 for Y_1 and 2.0 for Y_2. Usually it is hard to find the setting of the control factors at which both target values are met. In multiple response problems, we often seek to minimize a quadratic loss function, e.g. Khuri and Conlon ([22]), and Vining ([39]). As pointed by Khuri and Conlon ([22]), a natural optimization criterion can be defined as follows. In an m-dimensional multiresponse problem, let ϕ_k be the optimum values for the k^{th} response y_k, optimized individually over the experimental region, $k = 1, 2, \ldots, m$, and let $\Phi = (\phi_1, \phi_2, \ldots, \phi_m)'$. An operating condition x_0 is said to be optimal if

$$\mathbb{D}[\underline{\hat{y}}(x_0), \Phi] = [\underline{\hat{y}}(x_0) - \Phi]'\hat{\Sigma}_{(x_0)}^{-1}[\underline{\hat{y}}(x_0) - \Phi] \tag{5.3}$$

achieves its minimum at x_0 over the experimental region. The measure defined in (5.3) computes the weighted distance between the predicted mean responses and the target values, while the weight is scaled by the covariance matrix. While in dual response problems, we aim to minimize the variability as well as the difference between the predicted responses and the targets. Therefore, we develop another measure of performance based on 5.3:

$$\varphi[\underline{\hat{y}}(x_0), \Phi] = [\underline{\hat{y}}(x_0) - \Phi]'[\underline{\hat{y}}(x_0) - \Phi] + |\Sigma_{x_0}|. \tag{5.4}$$

The first part of (5.4) is the squared deviance from the target values based on the predicted mean responses, and the second part is the variability of the responses at that location. Therefore, minimizing the sum of the two parts is to reach a tradeoff between the deviance and the variance, which is analogous to the MSE criterion in the univariate case. With the above estimated models, we seek to minimize the function in (5.4) and find that at the location $x_0 = (-1.000, 0.9671, 1.000)$, $\varphi[\underline{\hat{y}}(x_0), \Phi]$ achieves its minimum value 204.976. At that location, the predicted responses $\hat{y}_1 = 992.92$ and $\hat{y}_2 = 2.4931$.

With the fitted models available, the practitioners are free to use any sensible optimization criterion to identify an optimal setting. Moreover, the Bayesian analysis provides the posterior distributions of those model parameters. Therefore we can also use the proportion of conformance (Wang and Lam [42]) to maximize the probability that both responses simultaneously meet their respective specification limits.

5.4 Comparison of Modelling Efficiencies Under Covariance Matrix Uncertainty

In this section we use simulated data to compare the Bayesian modelling approach using noninformative priors with Chiao and Hamada's modelling approach (referred as C&H henceforward) under different correlation and variance scenarios. Bivariate dual response data are simulated, though simulation studies can be easily extended to multivariate cases. The simulation setting is described below:

- There are 3 control factors which have been identified as important for the multiple responses included in the experiment;
- A CCD with the axial distance $\alpha = 1.682$ is used;

- 3 replicates are taken for each response at each design point;
- The variance follows a Log-Normal distribution;
- The models of the mean, variance and correlation are assumed to be known in advance, so there is no model misspecification problem involved;
- The mean model is always of full second order, and the correlation model only contains linear terms;
- Three different variance models are considered: linear terms only, linear terms and interactions, and complete second order models; and
- Three different correlation scenarios are considered: the correlation is dependent on the setting of the control factors, high correlation ($\rho = 0.9$), and low correlation ($\rho = 0.001$).

The above simulation is planned mainly to study the modelling performance under different covariance matrices in the most ideal and simplest case: no model misspecification, and plenty of information provided for the estimation of correlation models.

To compare the modelling efficiency, we extend the definition of the relative efficiency measure in the univariate dual response surface case. In the univariate case, the relative efficiency at the location x_0 is the ratio of the true minimum MSE against the predicted MSE at the location x_0. In the multivariate case, we can use the optimization criterion proposed in (5.4) as the multivariate version of the MSE criterion. Then the relative efficiency in the multivariate case can be defined as:

$$\text{relative efficiency} = \frac{\min \varphi_{\text{true}}}{\varphi_{\text{point proposed}}}.$$

Table 5.2 lists the relative efficiencies of the two modelling methods for different variances and correlations. Within each correlation scenario, as the number of variables contained in the variance model increases, the relative efficiencies decrease (the means and

Table 5.2: Comparison of the relative efficiencies under different correlations. The simulation size is determined such that the larger standard error of the relative efficiencies for the two methods does not exceed 0.01.

Correlation model	**Variance model**	**C&H**			**BAYES**		
		mean	**median**	**stdev**	**mean**	**median**	**stdev**
location dependent	linear	0.7039	0.7102	0.1722	0.8434	0.9091	0.0961
	linear+inter	0.6895	0.7614	0.2585	0.8520	0.9051	0.1547
	full second	0.5090	0.4780	0.2542	0.6922	0.7414	0.2269
$\rho = 0.9$	linear	0.7642	0.8246	0.1865	0.8266	0.8729	0.1461
	linear+inter	0.5010	0.4738	0.1815	0.8279	0.8761	0.1544
	full second	0.3820	0.3532	0.1960	0.5745	0.5539	0.2280
$\rho = 0.001$	linear	0.7099	0.8095	0.2575	0.8540	0.9074	0.1533
	linear+inter	0.5554	0.5886	0.2484	0.7120	0.7631	0.2230
	full second	0.5219	0.5076	0.2584	0.7448	0.8030	0.2023

medians get smaller) and become unstable (standard deviations get larger). This is expected, as the amount of information provided by the data does not change, while the model to be estimated becomes larger.

When the variance model contains only the linear terms, the performances of the two methods are similar for different correlations. However, when the variance model becomes complicated, their performances degrade at different levels when the correlation is fixed at 0.9 or 0.001. One explanation is that the complicated variance model accentuates the inappropriateness of over-parameterizing the correlation model. When the correlation is 0.9, the two responses are highly correlated. Under that condition, we should either drop one of them or use dimension reduction methods to represent the two responses with one characteristic (Khuri and Conlon [22]) before proceeding further. When the correlation is as low as 0.001, the two responses are almost independent, and then the analysis can be simplified into two univariate dual response problems. Exerting a correlation model into the analysis is over-parameterization. In both cases, the average relative efficiencies keep above 0.7 except when the variance model is of second order and the correlation is 0.9. Overall, the BAYES method is always more efficient than the C&H method as the BAYES method takes the non-constant covariance matrix into account.

The above analysis displays some preliminary work on the Bayesian hierarchical model for multivariate dual response surfaces. There is still a lot to be done. First, the propriety of the posterior distribution needs to be shown if noninformative priors are used for those hyperparameters. Second, the comparison between the Bayesian method and the C&H method can be carried out under more different situations, such as varying the design strategies, the number of control factors, and the class of the data distributions.

Chapter 6

Summary and Future Research

The primary goal of this research is to develop and study the Bayesian hierarchical approach in modelling dual response surfaces. The performance of the Bayesian model is evaluated relative to traditional least squares methods under different scenarios. Some preliminary research has shown that the application of the Bayesian method can be extended to multivariate dual response surfaces. Additional research has suggested that the Genetic Algorithm can be very efficient and straightforward in dual response surface optimization.

In Chapters 2 and 3, the Bayesian hierarchical approach is developed for the models with fully and partially replicated dual response designs, respectively. The associated theoretical and computational issues are discussed. Data sets from the literature are used to illustrate the application of the Bayesian method and subsequent optimization procedures based on the Bayesian inference.

Chapter 4 contains the results of an evaluation of the performance of the Bayesian approach relative to two least squares methods. These results show that under different design strategies, experimental settings, variance scenarios and data distributions, the Bayesian method is overall more efficient and robust.

Chapter 5 extends the application of the Bayesian method to multivariate dual response surfaces. In multivariate cases, the Bayesian method can incorporate the non-constant covariance matrices into the analysis in a more natural way. Some theoretical issues need to be investigated further for the multivariate cases.

Up to now, the research is restricted to using noninformative priors in the absence of subjective information. Noninformative priors put no influence upon the analysis so that the information from the data dominates the results. Usually, Bayesian analysis will show more advantages if relevant subjective information can help analyze the current data.

If there are some sequential designs which allow us to learn about the mean and variance response surfaces step by step, there will be historical information about those parameters when the final stage data is used to establish the dual models and search for the optimum points. The requirement for sequential designs is very reasonable in RSM, since most RSM are sequential in nature (e.g see Myers and Montgomery [26]). Usually after the screening experiment stage and after the design region has been moved to a near optimal area by using the method of steepest ascent/descent or some other methods, much valuable information has been obtained and informative priors are ready to be imposed on the parameters. However, eliciting informative priors is not easy. There are several issues to be investigated: how to represent historical data with some prior distributions; how to adjust the next stage designs according to the prior information; and how to assign the weight of the information from different historical periods into the current data analysis. All of these can be studied in the future.

Bibliography

[1] Aitkin, M. Modelling variance heterogeneity in normal regression using GLIM. *Applied Statistics 36* (1987), 332–339.

[2] Ames, A. E., Mattucci, N., Szonyi, G., and Hawkins, D. M. Quality loss function for optimization across mulitple response surfaces. *Journal of Quality Technology 29* (1997), 339–346.

[3] Bartlett, M. S., and Kendall, D. G. The statistical analysis of variance heterogeneity and the logarithmic transformation. *Journal of the Royal Statistical Society B8* (1946), 128–150.

[4] Box, G. E., and Draper, N. R. *Empirical Model-Building and Response Surfaces.* John Wiley & Sons, New York, 1987.

[5] Box, G. E. P., and Wilson, K. B. On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society B 13* (1951), 1–45.

[6] Chen, Y., and Ye, K. A Bayesian hierarchical approach to dual response surface modelling. Submitted.

[7] Chen, Y., and Ye, K. Bayesian hierarchical modelling on dual response surfaces in partially replicated designs. In preparation.

[8] Chiao, C. H., and Hamada, M. Analyzing experiments with correlated multiple responses. *Journal of Quality Technology 33* (2001), 451–465.

[9] Copeland, K. A. F., and Nelson, P. R. Dual response optimization via direct function minimization. *Journal of Quality Technology 28* (1996), 331–336.

[10] De Jong, K. A. *An Analysis of the Behavior of a Class of Genetic Adaptive Systems.* PhD thesis, 1975.

[11] Del Castillo, E. Multiresponse process optimization via constrained confidence regions. *Journal of Quality Technology 28* (1996), 61–70.

[12] Del Castillo, E., and D.C., M. A nonlinear programming solution to the dual response problem. *Journal of Quality Technology 25* (1993), 199–204.

[13] Derringer, G., and Suich, R. Simultaneous optimization of several response variables. *Journal of Quality Technology 12* (1980), 214–219.

[14] Engel, J., and Huele, A. F. A generalized linear modelling approach to robust design. *Technometrics 38* (1996), 365–373.

[15] Fletcher, R., and Powell, M. J. D. A rapidly convergent descent method for minimization. *Computer Journal 6* (1963), 163–166.

[16] Gelman, A., Carlin, J. B., Stern, S. H., and Rubin, B. D. *Bayesian Data Analysis.* Chapman & Hall/CRC, 2002.

[17] Hartley, H. O. Smallest composite designs for quadratic response surfaces. *Biometrics 15* (1959), 611–624.

[18] Hoke, A. T. Economical second-order designs based on irregular fractions of the 3^n factorial. *Technometrics 16* (1974), 375–384.

[19] Holland, J. H. *Adaptation in Natural and Artificial Systems.* Ann Arbor, MI: University of Michigan Press, 1975.

[20] Horn, R. A., and Johnson, C. R. *Matrix Analysis.* Cambridge: Cambridge University Press, 1985.

[21] I., K. A., and Cornell, J. A. *Response Surfaces: Designs and Analyses.* Marcel-Dekker, New York, 1996.

[22] Khuri, A. I., and Conlon, M. Simultaneous optimization of multiple response represented by polynomial regression function. *Technometrics 23* (1981), 363–375.

[23] Kim, K. J., and Lin, D. K. J. Dual response surface optimization: A fuzzy modelling approach. *Journal of Quality Technology 30* (1995), 1–10.

[24] Lin, D. k. J., and Tu, W. Dual response surface optimization. *Journal of Quality Technology 27* (1995), 34–39.

[25] Luner, J. J. Achieving continuous improvement with the dual response approach: A demonstration of the roman catapult. *Quality Engineering 6* (1994), 691–705.

[26] Myers, R. H., and C, M. D. *Response Surface Methodology: Process and Product Optimizaiton Using Design Experiments.* John Wileys & Sons, Inc, 2002.

[27] Myers, R. H., and Carter, W. H. Response surface techniques for dual response systems. *Technometrics 15* (1973), 301–317.

[28] Nair, V. N., Abraham, B., MacKay, J., Box, G., Kacker, R., Lorenzen, T., Lucas, J., Myers, R., Vining, G., Nelder, J., Phadke, M., Sacks, J., Welch, W., Shoemaker, A., Tsui, K., Taguchi, S., and Wu, C. F. J. Taguchi's parameter design: A panel discussion. *Technometrics 34* (1992), 127–161.

[29] Notz, W. Minimal point second order design. *Journal of Statistical Planning and Inference 6* (1982), 47–58.

[30] Peterson, J. J. A posterior predictive approach to multiple response surface optimization. *Journal of Quality Technology 36* (2004), 139–153.

[31] Pignatiello, J. J., J. Strategies for robust multiresponse quality engineering. *IIE Transactions 25* (1993), 5–15.

[32] Rao, C. R. *Linear Statistical Inference and Its Application.* John Wiley & Sons, 1973.

[33] Robert, C. P., and Casella, G. *Monte Carlo Statistical Methods*. Springer-Verlag, 1985.

[34] Shoemaker, A. C., Tsui, K.-L., and Jeff Wu, C. F. Economical experimentation methods for robust design. *Technometrics 33* (1991), 415–427.

[35] Srinivas, M., and Patnaik, L. Adaptive probabilities of crossover and mutation in genetic algorithms. *IEEE Transactions on Systems, Man and Cybernetics 24* (1994).

[36] Taguchi, G. Introduction to quality engineering: designing quality into products and processes. Tech. rep., Asian Productivity Organization, Tokyo, 1985.

[37] Tang, L. C., and Xu, K. A. A unified approach for dual response optimization. *Journal of Quality Technology 34* (2002), 437–447.

[38] Tong, L., and Su, C. Optimizing multiresponse problems in the Taguchi method by fuzzy multiple attribute decision making. *Quality and Reliability Engineering International 13* (1997), 25–34.

[39] Vining, G. G. A compromise approach to multiresponse optimization. *Journal of Quality Technology 30* (1998), 309–313.

[40] Vining, G. G., and Myers, R. H. Combining taguchi and response surface philosophies: A dual response surface approach. *Journal of Quality Technology 22* (1990), 38–45.

[41] Vining, G. G., and Schaub, D. A. Experimental strategies for estimating mean and variance function. *Journal of Quality Technology 28* (1996), 135–147.

[42] Wang, C. M., and Lam, C. T. Confidence limits for proportion of conformance. *Journal of Quality Technology 28* (1996), 439–445.

[43] Welch, W. J., Tat-Kwan, Y., Kand, S. M., and Sacks, J. Computer experiments for quality control by parameter design. *Journal of Quality Technology 22* (1990), 15–22.

Vita

Younan Chen was born on April 4, 1978 to Zhiliang Chen and Qiu'er Zhao in Zhoushan, Zhejiang province, China. She graduated from Putuo High School in Zhejiang province in 1997. In July 2001 she graduated from Beijing Foreign Studies University in Beijing, China with a B.A. degree in Economics. The author then started graduate studies at Virginia Polytechnic Institute and State University. She was awarded the M.S. degree in Statistics in December, 2002. After that, she began to pursue a Ph.D. degree in Statistics at Virginia Polytechnic Institute and State University, which she received three years later.

The author is a member of the honor society Mu Sigma Rho. She is also a member of the American Statistical Association.